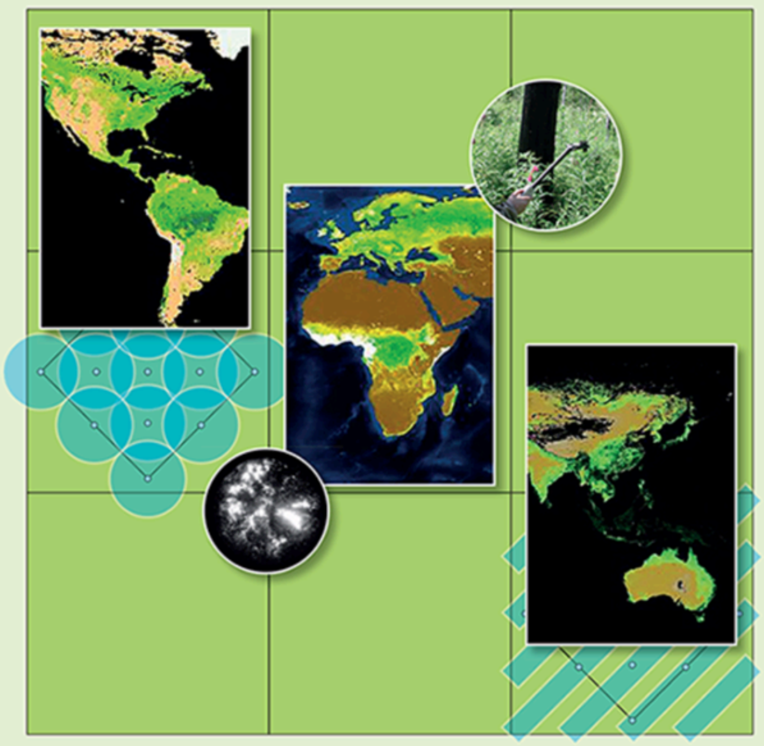
Citation:
Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., LeBlanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., & Widlowski, J.L. (2014). Global Leaf Area Index Product Validation Good Practices. Version 2.0. In G. Schaepman-Strub, M. Román, & J. Nickeson (Eds.), Best Practice for Satellite-Derived Land Product Validation (p. 76): Land Product Validation Subgroup (WGCV/CEOS), doi:10.5067/doc/ceoswgcv/lpv/lai.002

Committee on Earth Observation Satellites
Working Group on Calibration and Validation

Land Product Validation Sub-Group

Global Leaf Area Index Product Validation Good Practices



Editors: Gabriela Schaepman-Strub, Miguel Roman, Jaime Nickeson

Authors: Richard Fernandes, Stephen Plummer, Joanne Nightingale, Fred Baret, Fernando Camacho, Hongliang Fang, Sebastien Garrigues, Nadine Gobron, Matt Lang, Roselyn Lacaze, Sylvain LeBlanc, Michele Meroni, Beatriz Martinez, Tiit Nilson, Bernard Pinty, Jan Pisek, Oliver Sonnentag, Alexander Verger, Jon Welles, Marie Weiss, Jean-Luc Widlowski

Version 2.0.1
August, 2014

List of Revisions

Version	Revision	Date	Author
V1.0	Initial draft sent for peer review	February 2012	Fernandes and Plummer
V2.0	Peer review comments incorporated. Public version made available on LPV website	January 2014	Fernandes, Plummer and Nightingale
V2.0.1	Edited table 3. Added links, updated caption.	May 2014	Nickeson
V2.0.1	Added Appendix A	August 2014	Nickeson

Citation:
Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., LeBlanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., & Widlowski, J.L. (2014). Global Leaf Area Index Product Validation Good Practices. Version 2.0. In G. Schaepman-Strub, M. Román, & J. Nickeson (Eds.), *Best Practice for Satellite-Derived Land Product Validation* (p. 76): Land Product Validation Subgroup (WGCV/CEOS), doi:10.5067/doc/ceoswgcv/lpv/lai.002

Editor's Note

This document represents the views of the biophysical product focus area of the CEOS WGCV Land Product Validation (LPV) sub-group. This focus area provides those involved in producing and validating satellite based leaf area index (LAI) products with a forum for documenting accepted best practices in an open and transparent manner that is scientifically defensible. This Global LAI product validation best practice protocol document (V2.0) has undergone scientific review by remote sensing experts from across the world. All comments and suggestions have been considered to formulate this consensus document and responses to reviewer concerns are logged alongside the protocol on the LPV webpage. Furthermore, a list of recommendations arising from findings in this document will be provided on the LPV webpage (http://lpvs.gsfc.nasa.gov/). It is expected the best practice protocol document and recommendations will undergo subsequent regular iterations based on community feedback and scientific advancement.

We welcome all interested experts to participate in improving this document and invite the broader community to make use of it for their research and applications related to leaf area index products derived from satellite imagery. All contributors will be recognised as such in the document and on the CEOS WGCV LPV website.

Sincerely,

Richard Fernandes, Canada Centre for Remote Sensing
Stephen Plummer, European Space Agency
Joanne Nightingale, National Physical Laboratory

Chair of the CEOS WGCV Land Product Validation Group
Gabriela Schaepman-Strub, University of Zurich

January 2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUMMARY

The Global Climate Observing System (GCOS) has specified the need to systematically produce and validate global leaf area index (LAI) products. This document provides recommendations on good practices for the validation of global LAI products. Internationally accepted definitions of LAI and associated quantities are provided to ensure thematic compatibility across products and reference datasets. A survey of current validation capacity indicates that progress is being made towards the use of standard spatial sampling and in situ measurement methods, but there is less standardisation with respect to performing and reporting statistically robust comparisons. Three comparison approaches are identified: direct validation, indirect validation, and completeness. Direct validation, corresponds to the comparison of temporally and spatially concurrent satellite-derived product and up-scaled in situ reference LAI estimates. Indirect validation, consisting of inter-comparisons of products with ensembles of other products, using a stratified spatial sampling is proposed as a means for quantifying product precision as well as the representativeness of direct validation sites for a given biome. Completeness, corresponding to the frequency and continuity of LAI products, is quantified using a standard set of metrics applied to multi-year products. Finally, the need for an open access facility for performing validation as well as accessing reference LAI maps and ensemble LAI estimates from products is identified.

LIST OF ACRONYMS AND NOMENCLATURE

AAFC	Agriculture and Agrifood Canada
BIGFOOT	A NASA funded project linking in situ measurements, remote sensing, and models to validate MODIS products related to the terrestrial carbon cycle
BELMANIP(2)	CEOS WGCV Global Stratification for LAI Validation. See (Baret *et al.* 2006) for V1. V2 is a revisit of V1 to make it more compatible with the needs of validation and inter-comparison of 1km products
BOREAS	Boreal Ecosystem-Atmosphere Study (NASA)
BRDF	Bi-directional Reflectance Distribution Function
CANEYE	Imaging software used to extract canopy structure characteristics from true-colour images (either acquired with a fish-eye or with a long focal length objective lens)
CCRS	Canadian Center for Remote Sensing
CEOS	Committee on Earth Observation Satellites
CONECOFOR	European forest monitoring network
CSIRO	Commonwealth Scientific and Industrial Research Organization (Australia)
CYCLOPES	European project on biophysical parameter mapping.
DECAGON	Manufacturer of LAI survey instruments
DHP	Digital Hemispherical Photograph
DP	Long focal length digital photographs
DSLR	Digital Single Lens Reflex
DUE	Data User Element (ESA)
ECV	Essential Climate Variable (GCOS)
EPIFOV	Effective Projected Instantaneous Field of View
ESA	European Space Agency
ESU	Elementary Sampling Unit
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FLUXNET	Global network of flux tower sites.
FOREMON	European forest monitoring network
FOV	Field of View
FUTMON	European forest monitoring network
GCM	General Circulation Model
GCOS	Global Climate Observing System
GEOLAND	A European Union project to develop and demonstrate a range of reliable, affordable and cost efficient European geo-information services for land under the GMES Programme
GEOV1	Geoland2 Version 1 products
GLA	Gap Light Analyzer software for processing DHP
GLOBCARBON	An ESA initiative to generate multi-sensor estimation of global biophysical products for global terrestrial carbon studies
GMES	Global Monitoring for Environment and Security
GOFC GOLD	Global Observations of Forest and Land Cover Dynamics
GTOS	Global Terrestrial Observing System
HEMISFER	Software for LAI estimation from DHP

ICOS	Integrated Carbon Observing System: A European infrastructure dedicated to high precision monitoring of greenhouse gas fluxes
INRA	Institut National de la Recherche Agronomique (France)
IPCC	Inter-governmental Panel on Climate Change
ISO	International Organization for Standardisation
JRC	Joint Research Centre (European Union)
LAD	Leaf Area Distribution
LAI	Leaf Area Index
LAIe	Effective Leaf Area Index
LAINet	Network of LAI monitoring sites across the U.S.A.
Landsat ETM+	Landsat Enhanced Thematic Mapper +
Landsat TM	Landsat Thematic Mapper
LICOR	Company producing LAI survey equipment.
LIDAR	Light Detection and Ranging
LPV	Land Product Validation (sub-group of CEOS WGCV)
LTER	Long Term Ecological Research Network
MERIS	MEdium Resolution Imaging Spectrometer
MISR	Multi-angle Imaging Spectro-Radiometer
MODAPS	The MODIS Adaptive Processing System generates Aqua and Terra L1, atmosphere, and land data products on a near-real time basis using L0, ancillary, and ephemeris and attitude data products
MODIS	Moderate Resolution Imaging Spectro-radiometer (NASA)
NASA	National Aeronautics and Space Administration (USA)
NEON	National Environmental Observation Network (USA)
NPV	Non-Photosynthetic Vegetation
OLIVE	On-Line Validation Exercise
ORNL	Oak Ridge National Laboratory (USA)
PAI	Plant Area Index
PAIe	Effective Plant Area Index
PEN Japan	Phenological Eyes Network of Japan
PIFOV	Ground Projected Instantaneous Field of View
POLDER	Polarization and Directionality of the Earth's Reflectances – Satellite sensor
QA	Quality Assessment
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SURFRAD	Network of surface radiation measurement sites
Tartu	Tartu Observatory Leaf Area Index Survey Protocol
TIP	JRC Inversion Package for Biophysical Parameter Estimation
TOA	Top of Atmosphere
TRAC	Tracing Radiation and Architecture in Canopies
UNFCCC	United Nations Framework Convention on Climate Change
USA NPN	USA National Phenological Network
VALERI	Validation of Remote Sensing Instruments LAI reference map protocols

VGT	The SPOT-Vegetation imager on SPOT4 and 5 satellites
WGCV	Working Group on Calibration and Validation (CEOS)
WMO	World Meteorological Organization

1 INTRODUCTION

This section explains the international framework that has motivated the current document, describes LAI requirements within this framework and summaries the goals of this document.

1.1 Importance of LAI

Leaf Area Index (LAI) measures the amount of plant leaf material in an ecosystem. It appears in many models describing vegetation-atmosphere interactions (GCOS-138 2010) as a key variable controlling processes such as photosynthesis, respiration and rain interception. It is defined as one half the total green leaf area per unit ground surface area (Chen *et al.* 1992). On sloping surfaces, the LAI should be projected to the normal to the slope.

1.2 The UNFCCC and the Global Climate Observing System

Worldwide systematic observation of the climate system is a key prerequisite for advancing scientific knowledge on climate change. The United Nations Framework Convention on Climate Change (UNFCCC) calls on Parties to promote and cooperate in systematic observation of the climate system, including through support to existing international programmes and networks, as indicated in Articles 4.1(g) and 5 of the Convention. A key dimension for the implementation of those Articles has been the cooperation with the Global Climate Observing System (GCOS), a joint undertaking of the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational Scientific and Cultural Organization (UNESCO), the United Nations Environment Programme (UNEP) and the International Council for Science (ICSU) with its secretariat hosted by the WMO, reinforced by decisions taken at various Conferences of the Parties. The signatories of the UNFCCC have thus adopted the GCOS as the organising body for climate observations expressed through its Implementation Plans (GCOS-92 2004, GCOS-138 2010). These Implementation Plans establish the requirements for the systematic monitoring of a suite of Essential Climate Variables (ECVs) globally. Leaf Area Index (LAI) is one of the *16* terrestrial ECVs (GCOS-138 2010).

1.3 The Role of CEOS WGCV

LAI can be measured in situ (see Section 3.2) and indirectly from space-based observations. While it is routinely measured at a number of research sites, the measurement network is sparse in many regions of the world. The CEOS Cal/Val Portal currently hosts in situ reference LAI data from 113 global direct validation sites shown in Figure 1 (http://calvalportal.ceos.org/web/olive/site-description). This network should be maintained and ideally expanded to become much more representative of the diversity of global biomes and ecosystem conditions. The CEOS Cal/Val Portal also identifies a globally representative sampling (BELMANIP2) that should be targeted for future direct validation sites.

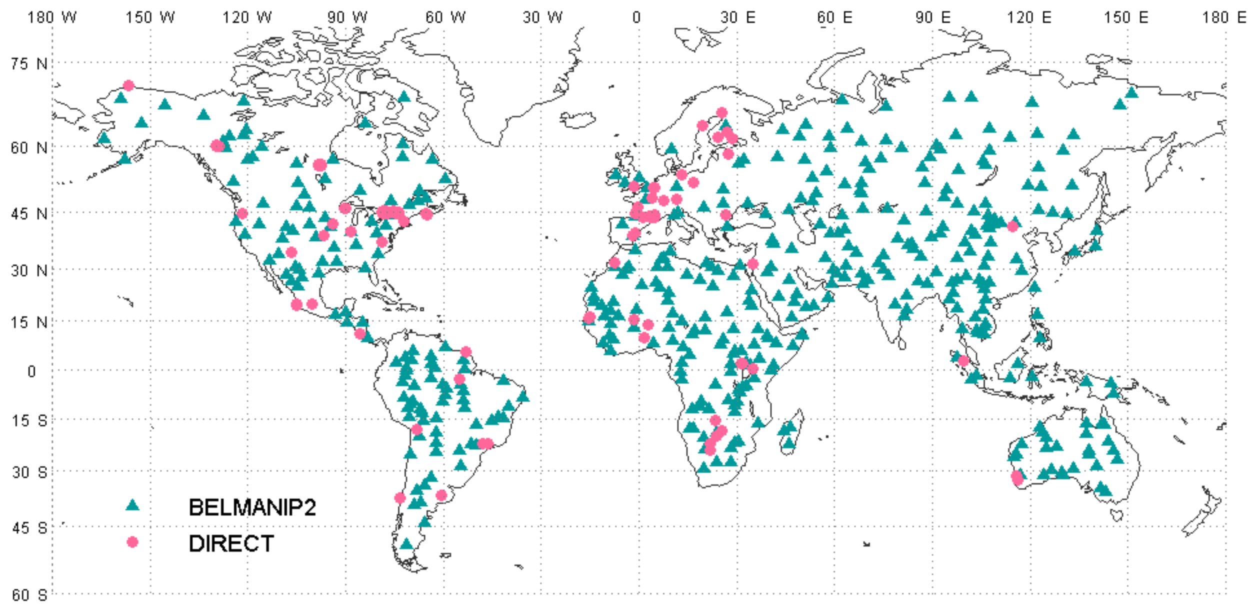


Figure 1: Location of reference LAI sites available for direct validation and BELMANIP2 sites designated for product inter-comparison based on the OLIVE Validation Platform. (http://calvalportal.ceos.org/web/olive/site-description).

The process of improving both the space-based observations and the in situ network is embodied in the GCOS Implementation Plans and the accompanying Satellite Supplements (GCOS-107 2006, NEON 2009). The Committee on Earth Observation Satellites (CEOS) Working Group on Calibration and Validation (WGCV), and in particular its sub-group on Land Product Validation (LPV), are required to play a key coordination role as well as lend the expertise required to address actions related to validation of global LAI measurements as identified in GCOS-138 as follows:

a. LAI can be estimated in situ by destructive sampling or with the help of commercially available dedicated instruments. It is routinely measured at a number of research sites dealing with surface climate, ecological, or agricultural issues. CEOS WGCV is playing a coordinating role in this work. Benchmarking and consistency checking are required for the global archive of LAI measurements (p124 of GCOS-138).

b. The development and maintenance of reference sites to address [the] inadequacy [in the reference network in some parts of the globe] should be addressed [….]. Building on existing networks, such as FLUXNET, LAInet and BIGFOOT, is a possible way to improve this situation. The CEOS WGCV has begun to coordinate this through the creation of a centralised database, an activity that should continue (p124).

c. Benchmarking and comparison of [satellite] LAI products is essential to resolve differences between products and to ensure their accuracy and reliability. The CEOS WGCV should lead this activity in collaboration with GCOS and GTOS, exploiting in situ observations from designated reference sites and building on the validation activities currently being undertaken by the space agencies and associated research programmes (p125).

CEOS considers these roles important to achieving validated global LAI products, but at the same time recognises current limitations in both resources and in some cases knowledge within both CEOS and the international expert community. This good practice document makes recommendations that if followed should serve to address many of the current limitations.

1.4 GCOS IP Action Items

The role of the CEOS WGCV has been consolidated in a series of Action Items in the GCOS Implementation Plan (GCOS-138 2010):

In the terrestrial domain it is essential to obtain global products for most ECVs from a range of satellite sensors supported by in situ measurements. A coordinated in situ network of terrestrial reference sites is needed for (p14):

a. Observations of the fullest possible range of terrestrial ECVs and associated details relevant to their application in model validation;
b. Process studies;
c. Validation of observations derived from Earth observation satellites; and
d. To address intrinsic limitations in some of these, such as the saturation of LAI measurements.

Listed below are three key requirements for in situ measurements at reference sites in the context of long-term global climate measurements (p106):

a. To ensure that a representative set of biomes are properly and consistently documented over long periods of time (decades or more). This will allow the details of natural vegetation changes and carbon stocks, including fluxes, to be carefully monitored at key locations.
b. To measure key meteorological ECVs to support interpretation of changes recorded at such sites.
c. To optimise the joint use of these terrestrial reference sites with:

   i. A set of sites delivering essential ground data for the validation of satellite-derived products that provide extensive geographical coverage for these variables (see Action T29 dealing specifically with calibration/validation of FAPAR and LAI).

ii. A set of key ecosystem sites (see Action T4).

Currently available satellite LAI products have been shown to exhibit significant differences (Garrigues *et al.* 2008a, Camacho *et al.* 2010, Camacho *et al.* 2011, Fang *et al.* 2013), which detract from their usefulness in downstream applications (Gobron *et al.* 2009).

In responding to GCOS, CEOS has assigned the action items T29 and T30 to the Biophysical Focus Area of the Land Product Validation Sub-group of the CEOS WGCV (LPV-Biophysical). LPV-Biophysical has submitted a proposal for a calibration/validation in situ network to GCOS (http://lpvs.gsfc.nasa.gov/LAI_background.html) in response to T29. This document contributes to addressing the action item T30.

Action T29 *[IP-04 T29]*98 **Action:** Establish a calibration/validation network of *in situ* reference sites for FAPAR and LAI and conduct systematic, comprehensive evaluation campaigns to understand and resolve differences between the products and increase their accuracy.
Who: Parties' national and regional research centres, in cooperation with space agencies coordinated by CEOS WGCV, GCOS and GTOS.
Time-Frame: Network operational by 2012.
Performance Indicator: Data available to analysis centres.
Annual Cost Implications: 1-10M US$ (40% in non-Annex-I Parties).

Action T30 *[IP-04 T30]* **Action:** Evaluate the various LAI satellite products and benchmark them against *in situ* measurements to arrive at an agreed operational product.
Who: Parties' national and regional research centres, in cooperation with space agencies and CEOS WGCV, TOPC, and GTOS.
Time-Frame: Benchmark by 2012.
Performance Indicator: Agreement on operational product.
Annual Cost Implications: 1-10M US$ (10% in non-Annex-I Parties).

1.5 LAI Requirements

LAI products are currently used over local (<10km^2), regional (<1000km^2) and global extents. Local and regional requirements vary significantly by intended use. However, GCOS has specified a set of global target requirements (GCOS-138 2010) that in many cases may meet local and regional needs:

Spatial resolution: 250 m horizontal
Temporal resolution: 2-weekly averages
Accuracy: maximum of 20% or 0.5
Stability: maximum of 10% or 0.25

GCOS has also specified a requirement for a near-term global LAI product (GCOS-138, Action Item 31) at 2km resolution or better although without specification of the temporal resolution.

The GCOS requirements are supplemented by application specific requirements identified by the WMO (Table 1). These specific requirements are defined at goal (ideal), breakthrough (optimum in terms of cost-benefit), and threshold (minimum acceptable). In most cases the GCOS requirements satisfy threshold levels (especially considering that GCOS requirements greatly exceed threshold spatial resolution requirements so random errors will cancel during spatial aggregation).

Table 1: WMO Requirements for Global LAI Products.
(From http://www.wmo-sat.info/oscar/variables/view/98); G=goal,B=breakthrough,T=threshold.

Application	Accuracy (%)			Spatial Resolution (km)			Temporal Resolution (d)		
	G	B	T	G	B	T	G	B	T
Global Weather Prediction	5	10	20	2	10	50	1	5	10
Regional Weather Prediction	5	10	20	1	5	40	.5	1	2
Hydrology	5	8	20	0.01	0.1	10	7	11	24
Agricultural Meteorology	5	7	10	0.01	0.1	10	5	6	7
Climate-Carbon Modelling	5	7	10	0.25	0.85	10	1	3	30

1.6 Goal of this Document

In response to GCOS Action Item T30, the goal of this document is to identify good practices for validating global satellite LAI products. The document will specifically address accuracy assessment against reference LAI measurements. The latter should be traceable to in situ measurements of known accuracy and the assessment augmented with metrics of precision derived from ensembles of products themselves.

2 DEFINITIONS

This section provides the necessary definitions relevant to global LAI validation.

2.1 Definition of LAI

LAI is defined as one half the total green leaf area per unit horizontal ground surface area (Chen *et al.* 1992, GCOS-138 2010). Green leaves correspond to vegetation matter capable of photosynthesis in ambient conditions. This definition was adopted across the various international groups (CEOS WGCV, GTOS, GOFC-GOLD, GCOS) in December 2010.This definition of LAI includes foliage in the overstory including epiphytes and foliage in the understory including mosses.

2.2 Definitions of Associated Physical Parameters

2.2.1 Projected LAI

Projected LAI corresponds to the projected area of all foliage in a region onto a plane normal to a specified direction. A commonly used normal is the vertical. In comparison to LAI, when using the vertical normal for projected LAI, if leaves were tilted away from the horizontal, the projected LAI will decrease with the lowest value for the vertically oriented leaves. Folding leaves and needle foliage will also modify the projected LAI in comparison to flat leaves.

Previously, both in situ measurements and satellite-derived products often reported projected LAI. Conversion factors are required if these data are to be used in a validation protocol. These factors can be derived using geometric information on foliage orientation e.g. (Barclay 1998) or by comparison to corresponding measurements of LAI. Ideally, both in situ reference LAI data and products should use LAI rather than projected LAI to avoid the complexity of developing these conversion factors and the additional uncertainty due to the factors when performing comparisons.

2.2.2 Plant Area Index (PAI)

PAI is half the total surface area of all above ground vegetation matter. Many in situ LAI estimates are based on indirect measurements related to gap fraction or transmission that cannot easily separate green leaf area from non-photosynthetic vegetation (NPV). By convention the NPV to total area index ratio ($\propto$) is used to relate LAI to PAI as:

$$LAI = (1-\propto)PAI \quad (1)$$

NPV includes standing and dead woody matter, leaf litter, dead moss or dead lichen and fruit but does not include green stalks and vegetation that is still capable of photosynthesis even if it is currently dormant (e.g. dried moss carpets).

2.2.3 Effective LAI (LAIe) or Effective PAI (PAIe)

LAI and PAI have frequently been estimated using in situ measurements of directional transmission of solar radiation or by gap fraction measurements from imaging sensors (Breda 2003, Jonckherre *et al.* 2004). These approaches are sensitive to the projected area of the foliage along each measurement direction and hence the selection of direction as well as the leaf angle distribution (Nilson 1971, Nilson 1999). Historically, two approaches have been used to minimise this sensitivity to foliage angle. The first (conventionally termed 'Miller's method' after (Miller 1967) is to measure the uncollided transmission or the gap fraction (P) along all zenith angles (θ) of the upper hemisphere (or lower hemisphere for gap fraction of low vegetation) and estimate the PAIe as:

$$PAIe = 2\int_{0}^{\frac{\pi}{2}} -\ln P(\theta)\cos\theta\sin\theta d\theta \qquad (2)$$

The second approach, termed here the '1 radian estimate' is to measure uncollided transmission or gap fraction at 1 radian (≈57°) from the normal to the local vertical datum so ensuring that the leaf projection coefficient of unit foliage area on a plane perpendicular to the view direction ("G-function", (Nilson 1971)) converges at approximately 0.5 irrespective of the leaf inclination angle distribution (Lang *et al.* 1986, Weiss *et al.* 2004):

$$PAIe = -0.92573\ln P(1) \qquad (3)$$

Both approaches are almost equivalent under ideal measurement conditions (Leblanc *et al.* 2005a).

The above theory for PAIe also applies to LAIe assuming that the canopy either has no NPV or that a correction factor relating PAIe to LAIe is applied. Further discussion of these terms is provided in (Breda 2003, Ryu *et al.* 2010).

2.2.4 Clumping Index

The clumping index is the ratio of the LAIe measured under conditions listed in Section 2.2.3 to the LAI (Nilson 1971, Chen *et al.* 1992):

$$\Omega = \frac{LAIe}{LAI} \qquad (4)$$

The same definition is often applied after replacing LAIe and LAI with PAIe and PAI respectively. This definition of clumping does not distinguish between different scales of clumping (e.g. crown, patches or rows of vegetation). Rather, it is provided here to note the requirement to include appropriate clumping index conversion factors when using reference LAIe measurements or when validating LAI products that are in fact calibrated to retrieve LAIe.

Under certain conditions the clumping index can be estimated using transmission measurements, gap fraction measurements or directional reflectance measurements (Chen *et al.* 1995, Leblanc *et al.* 2005a, Ryu *et al.* 2010). These estimates will be specific to the scale and spatial and directional sampling of measurements. For specific circumstances of needle leaf canopies an additional shoot clumping is needed to account for the non-random position of needles on shoots. Shoot clumping is specified as either the 'shillouette to total needle area ratio' STAR (Oker-Blom *et al.* 1988) or the needle to shoot area ratio (Chen *et al.* 2006). Values for needle to shoot area ratio range from 1.2 to 2.0 (Oker-Blom *et al.* 1988, Kucharik *et al.* 1998, Breda 2003, Chen *et al.* 2006).

2.3 Definition of Other Key Terms

Validation of satellite LAI products relies on aspects specific to satellite measurements. This section reviews terminology in this context.

2.3.1 Elementary Sampling Unit (ESU)

An Elementary Sampling Unit (ESU) is a contiguous spatial region over which the expected value of LAI can be estimated through in situ measurement. The ESU corresponds to the finest spatial scale of LAI estimates used for reference LAI maps. The ESU size is at least as large as one measurement footprint of the in situ instrument and typically includes a number of instrument measurements. The maximum ESU size is determined by the level of within ESU LAI variability that can be tolerated by the validation protocol and the effort available to conduct measurements. The size of each ESU within a reference region also varies with surface condition, instrument field of view, illumination conditions (when transmission based measurements are used) and spatial sampling design. For example, figure 2 indicates the sensitivity of the measurement field of view to canopy and illumination conditions for two common instruments.

The ESU size should be sufficient to allow repeat visit with minimum uncertainty due to changes in illumination or geolocation. Many indirect LAI estimates rely on statistical approaches that require a minimum spatial footprint, and hence ESU size, per measurement. Figures 3 and 4 show different ESU sizes determined by the combination of instrument, sampling design and canopy height. It is good practice to document the size of each ESU and relate the size to the measurement instrument, protocol and canopy height.

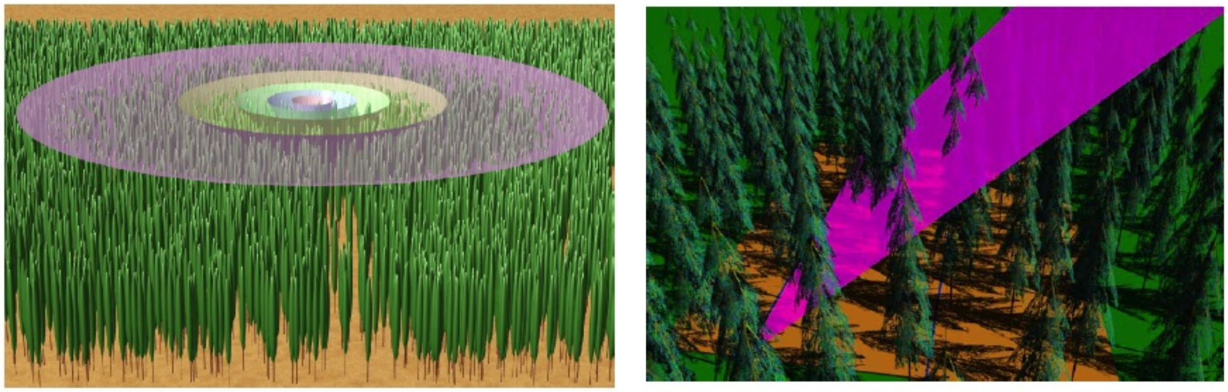


Figure 2: Depiction of spatial footprint of a LAI-2000 instrument as a function of zenithal view ring (left) and the TRAC instrument for a given solar illumination condition (from (Leblanc 2005c, Leblanc 2008).

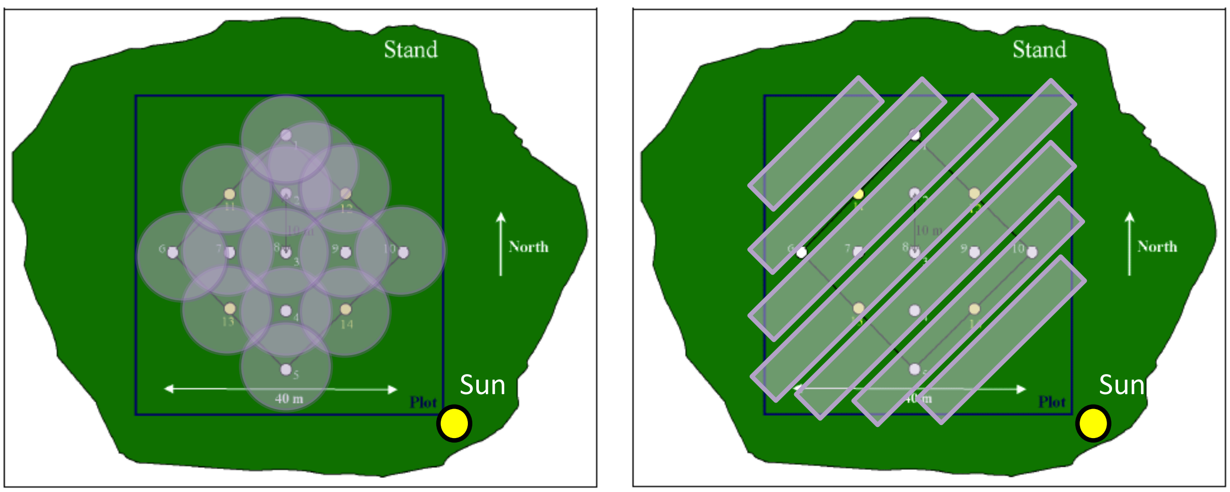


Figure 1: Spatial footprints of LAI-2000 (left) and TRAC (right) measurements following the CCRS sampling scheme (adapted from (Leblanc 2005c, Leblanc 2008)) for overstory LAI for a 40mx40m ESU. LAI-2000 footprints determined by canopy height while TRAC footprints are determined both by canopy height and solar zenith angle. Only every second TRAC footprint shown for clarity.

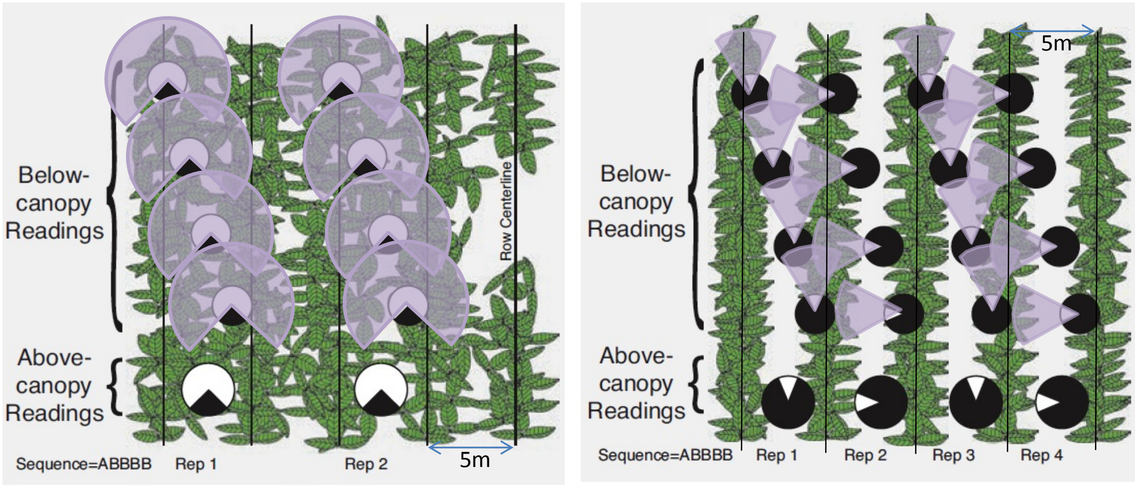


Figure 2: Spatial footprint (purple) of LAI-2000 sampling of homogenous (left) and row (right) 1m tall 20m x 20m crop canopy ESU following the LICOR protocol (adapted from (Licor 2010)). Note the large difference in sampled area with change in view cap to account for row crops. Normally this sampling would need to be replicated to cover an ESU of sufficient size for LAI validation purposes.

2.3.2 Local Horizontal Datum

The local horizontal datum is the plane containing the tangent to the local geoid corresponding to the centre of an ESU or mapping unit. For sloped terrain corrections to LAI or PAI estimates for the increased surface area of the slope may be required depending on survey method.

2.3.3 Ground Projected Instantaneous Field of View of Measurement (PIFOV)

The ground projected instantaneous field of view (PIFOV) is the area on the ground corresponding to the region over which a measurement is performed. For radiometric measurements, this area is defined as the region where the instrument point spread function, including all processing aspects except for spatial resampling, exceeds a specified threshold. The majority of imaging scanners including satellite imagers have PIFOV on flat ground on the order of twice the inter pixel sampling distance. In cases of off nadir measurements or large terrain slopes, the canopy height should be included when modelling the PIFOV for small footprint imaging scanner measurements. The PIFOV of an in situ instrument will vary with the canopy height and angular sampling of the instrument.

2.3.4 Effective Ground Projected Instantaneous Field of View of Measurement (EPIFOV)

The effective ground projected instantaneous field of view (EPIFOV) corresponds to the spatial extent of a measurement including both the PIFOV and the impact of spatial resampling. Resampling performed using smoothing filters (e.g. cubic convolution) will result in an EPIFOV on the order of the size of the PIFOV convolved with the size of the filter spatial support. Non-linear resampling, such as nearest neighbour can result in substantial spatial aliasing so that comparisons of values recorded in different EPIFOVs should include some sort of spatial averaging using a filter spatial support on the order of multiple PIFOVs.

2.3.5 Satellite Measurement Geolocation Uncertainty

Geolocation uncertainty, for LAI validation, corresponds to the planimetric uncertainty of a satellite measurement located on the same projection and datum as the ESU or study site reference LAI estimates. Geolocation uncertainty is often reported in nominal terms and based on a normal distribution of errors. Acquisition specific biases are often possible so that geolocation uncertainty should be visually assessed in comparison to reference vector layers whenever possible.

2.3.6 Mapping Unit

A mapping unit is the spatial region on the Earth's surface corresponding to a product or reference map value for a specified temporal extent. The majority of satellite based LAI products use mapping units corresponding to pixels within rasters in a specified map projection rather than per nominal EPIFOV location. As such, these products include a spatial generalisation corresponding to the transformation of the LAI estimate over each EPIFOV to the LAI estimate in the mapping unit. Considering that GCOS requires gridded LAI products at a constant spatial resolution, the CEOS LAI validation protocol assumes uncertainties due to this generalisation or due to temporal aggregation are considered in the total product uncertainty.

3 GENERAL CONSIDERATIONS FOR SATELLITE-DERIVED GLOBAL LAI PRODUCT VALIDATION

3.1 CEOS Validation Stages

The CEOS WGCV Land Product Validation sub-group has identified four validation levels corresponding to increasing spatial and temporal representativeness of samples used to perform direct validation (Table 2). The LAI validation protocol includes these aspects and supplements them with requirements for assessing the spatial and temporal precision of individual products.

Table 2: The CEOS WGCV Land Product Validation Hierarchy.

Stage 1 Validation	Product accuracy is assessed from a small (typically < 30) set of locations and time periods by comparison with in situ or other suitable reference data.
Stage 2 Validation	Product accuracy is estimated over a significant set of locations and time periods by comparison with reference in situ or other suitable reference data. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and time periods. Results are published in the peer-reviewed literature.
Stage 3 Validation	Uncertainties in the product and its associated structure are well quantified from comparison with reference in situ or other suitable reference data. Uncertainties are characterised in a statistically robust way over multiple locations and time periods representing global conditions. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and periods. Results are published in the peer-reviewed literature.
Stage 4 Validation	Validation results for stage 3 are systematically updated when new product versions are released and as the time-series expands.

3.2 Reference LAI Estimates

Reference LAI estimates are required to evaluate the accuracy and to a lesser extent the spatial and temporal precision of LAI products. These estimates can be derived by suitable up-scaling of either in situ LAI measurements over a number of ESUs or by acceptable qualitative estimates of LAI based on land cover. This section surveys approaches that have been used to perform these tasks and identifies good practices related to the production of reference LAI estimates.

3.2.1 ESU Mapping Unit

Most good practices for LAI validation require an estimate of the spatial mapping unit corresponding to each sampled ESU. The ESU mapping unit should correspond to the area over which the LAI together with its associated measurement error are representative. The ESU should also be large enough to be either directly co-located with LAI product mapping units (see section 5.1.3) or with ancillary information that can be used to upscale multiple ESUs over a region. The ESU should also include a means of estimation of measurement precision such as replicate sampling. These considerations often drive the specification of the ESU mapping unit.

Example 1: ESU Mapping Unit Specification - Destructive Sampling

Destructive sampling of all vegetation within ten $1m^2$ quadrats in a 10m diameter plot could provide LAI estimates for an ESU corresponding to each quadrat. In this case each ESU will have high measurement accuracy but the precision of each measurement is not easy to quantify since there is no replication. Quadrat level LAI information is not useful for direct comparison to global LAI products or even ancillary data (e.g. airborne imagery) unless the LAI within the ESU is spatially representative. Rather, the ESU should be defined as the 10m diameter plot at the expense of potentially increased measurement error. In this case the replicate sampling within the plot can be exploited to estimate measurement precision.

Example 2: ESU Mapping Unit Specification - LAI-2000 Row Crops

The 1m tall row crop surveyed in Figure 4 using the LAI-2000 falls within a 100m x 100m field with uniform planting and site conditions. The Licor manual suggests at least 23 measurements are required for a +/-10% precision for estimated LAI. To satisfy this requirement the sampling scheme shown in figure 4 should be extended along rows to include at least two more measurements per row (24 total). This would provide an ESU on the order of 20m x 20m within the field. Ideally a replicate ESU would be located in the field and surveyed by a different instrument and operator to quantify measurement error assuming within field LAI variation is small.

3.2.2 In situ Reference LAI over an ESU

In situ LAI estimates are acquired for diverse applications and by a range of approaches (Breda 2003, Jonckherre *et al.* 2004, Weiss *et al.* 2004, Leblanc 2005c, Law *et al.* 2008). Each estimate corresponds to the LAI of an ESU for a representative temporal interval. There is substantial debate as to the most appropriate measurement method and sampling e.g. (Garrigues *et al.* 2008b, Ryu *et al.* 2010). Robust methods in terms of ensuring reasonable precision have been developed and tested within existing survey networks (refer to table 3). All of these approaches are acceptable for the production of reference LAI maps provided that they are applied for the targeted land cover and that the ESU LAI uncertainty is also estimated. Recommendations for uncertainty estimation of ESU LAI are outlined within in Section 3.2.3.

Table 3: In-situ LAI guidelines.

Name	Destructive	Land Cover	Citation
AAFC	No	Short crops	Liu et al. 2013
BIGFOOT	Forest No Crops Yes	All	Cohen et al. 2006
BOREAS Destructive	Yes	Forest	Gower et al. 1997
BOREAS Non-Destructive	No	Forest	Chen et al. 1997
CCRS CANEYE	No	All	Fernandes, 2012
CCRS DHP	No	Forest, tall crops	Leblanc 2008
CCRS TRAC	No	All except short herbs	Leblanc 2005
CCRS Tundra	Yes	Short herbs	Chen et al. 2009
CONECOFOR	Yes	Deciduous forest	Cutini 2002
DECAGON Ceptometer	No	Short Herbs and crops	Decagon 2012
FLUXNET	Forest No Crops Yes	All except short herbs	Chen et. al. 2006
FUTMON	Both	Forest	Futmon, 2009
GTOS	Forest No Crops Yes	All except short herbs	Law et al. 2008
Helsinki University	No	Forests	Majasalmi et al. 2012
INRA Row Crop	No	Short crops and herbs	Baret et al.2010
LICOR LAI-2000/2200	No	All except short herbs	LICOR 2010
Ryu /Nilson	No	Forest	Ryu et al. 2010. Nilson and Kuusk, 2004
UNECE	Both	Deciduous forests	Pitman et al. 2010
VALERI	No	All	Weiss et al. 2004 Weiss 2002 Garrigues et al. 2006

This table lists current published methods for in-situ survey of leaf area index for an Elementary Sampling Unit. The column headers are: 'Name' which indicates the CEOS LAI CAL/VAL acronym for the method, hyperlinked to the guideline document, 'Destructive' indicates if the method involves destructive sampling as a primary requirement, ‘Land Cover’ indicates the land cover for which the method has or potentially could be applied, and ‘Citation’ indicates the reference for the method within the CEOS LPV Good Practices document. Note that conifer forests require correction for needle clumping, see Stenberg (1996) where the protocols do not provide details. See Appendix A for full URLs for each method.

Qualitative approaches can also be used to assign lower or upper bounds to LAI based on land cover and in situ datasets. For example, based on a global survey of in situ LAI (Asner *et al.* 2003) one can assign a minimum and maximum range of LAI as a function of land cover class (Table 4). This approach will have low accuracy in many instances but may be sufficient if the represented land cover class is a small proportion of the reference map.

Table 4: Range of LAI as a function of land cover (after (Asner *et al.* 2003)).

Land Cover Class	Minimum LAI	Maximum LAI	Restrictions
Crops	0.2	8.7	None
Desert	0.2	2.8	None
Forest – Needle leaf	0.01	15	None
Forest – Broadleaf	0.5	11.6	None
Grasslands - Prairie	0.3	5.0	North American only
Plantation	1.6	18.0	Not prior to leaf emergence
Shrubland	0.4	4.5	Not alpine or tundra
Grassland - Tundra	0.2	5.3	Tundra biomes

3.2.3 ESU LAI Accuracy

The statistical significance of differences between reference LAI maps based on ESU LAI and a satellite-derived LAI estimate will depend on the uncertainty of the reference maps and thus, to some extent, the uncertainty of the ESU LAI estimates. ESU LAI uncertainty should be reported as both an accuracy error and a precision error since the former may persist during inter-comparison with products, while the latter may cancel out across multiple ESUs.

Accuracy errors include best-case errors in the absence of measurement outliers and errors due to outliers in measurements. In the absence of local destructive sampling two good practices are recommended to quantify best-case errors.

1. An estimate of the expected value of best-case accuracy error for ESUs should be derived from median and extreme errors reported in literature studies relying on destructive sampling over ESUs of similar size (order of magnitude area), land cover and if possible species. This estimate will be pessimistic in that it will possibly include precision error. This estimate should include uncertainty in correction for canopy clumping and NPV area.

2. An independent estimate of the best-case accuracy error can be derived where feasible by comparing the LAI estimates over the same ESU using two different indirect measurement methods. This could mean the application of two methods using different angular sampling from the same instrument or two methods using two different instruments as long as their precision is comparable. This approach will tend to underestimate accuracy error in cases where methods share similar theoretical assumptions or correction factors for NPV. Nevertheless, it can serve

as a lower bound for accuracy error.

In the majority of cases the reported ESU LAI corresponds to the mean LAI over replicated measurements within the ESU. In practice replicate LAI measurements will also include noise due to operator error (e.g. poor pointing of instruments, biased classification of digital photographs) that may only apply to a few measurements. Good practices for designing measurement protocols for reducing these biases are beyond the scope of this document. However, if within ESU LAI measurements are available over some ESUs, the difference between the mean and median LAI can serve as an upper bound on this accuracy error, due to outliers, assuming at least 50% of the measurements are unbiased in the first place. It is good practice to report the total accuracy error for each ESU LAI estimate as the Euclidean sum of the larger of the two best-case error estimates and the outlier error.

Example 3: ESU LAI Accuracy

Assume Example 2 corresponds to a soybean crop with a mean LAI of 3.5 and median LAI of 2.5. The uncertainty of the LAI-2000 over soybean is on the order of ~0.5 based on comparison to destructive sampling (Welles *et al.* 1991, Malone *et al.* 2002) and ~0.5 units based on comparison to other instruments (Garrigues *et al.* 2008a). In this case the difference between mean and median LAI serves as a worst case accuracy error of 1.0 unit and the reported accuracy of 1.11 is the Euclidean sum of 0.5 and 1.0.

3.2.4 ESU LAI Precision

In terms of LAI product validation, ESU LAI precision is not as important as accuracy since spatially random ESU LAI errors will tend to cancel if unbiased spatial scaling methods are applied to estimate reference LAI maps. Nevertheless, LAI precision may be required as an input to these approaches or when ESU LAI is directly compared to LAI products.

ESU LAI precision can be quantified using repeat sampling of measurements within the ESU using the same instrument type, protocol and canopy conditions. Precision error for the estimate of ESU LAI will decrease as the number of measurements increase and as the natural variation of LAI decreases (e.g. one can expect high precision over a uniform LAI region with even a few measurements). It is good practice to report the 95th percentile interval for the ESU LAI estimate as a statistic related to LAI precision. A formula providing a good approximation for relative precision error δ corresponding to the 95%th percentile confidence interval of the ESU LAI is given in by (Licor 2010), Chapter 9)) and summarised here.

$$\delta = \frac{t(n)D}{LAI\sqrt{n}} \qquad (5)$$

Where, n is the number of repeat measurements, D is the standard deviation of the LAI within the ESU in the absence of measurement error and $t(n)$ is an approximation to the Student’s t distribution for a 5% probability:

$$t(n) = \frac{1}{0.060798n - 0.11528} + 2.9817 \quad (6)$$

The standard deviation of LAI, D, should be approximated as the standard deviation over the measurements although this will be an overestimate as it includes measurement error.
Repeat measurements in each ESU should be located independent of the actual pattern of vegetation in the plot and in a manner to minimise overlap between measurements. This may be accomplished by random sampling when the ratio of measurement sampling footprint to plot size is small or by a regular sampling that spaces footprints to minimise overlap when the ratio of measurement sampling footprint to plot size is small. Since regular sampling satisfies both requirements it is generally accepted as a good practice in most ESU LAI protocols as long as there is some element of randomness in the position of the centre of each measurement and of the sampling pattern to ensure sampling is not biased towards accessible gaps in the vegetation.

Example 4: ESU LAI Precision

Destructive sampling of all vegetation over ten $1m^2$ quadrats randomly sampled in a 10m diameter ESU gives a mean LAI of 4.0 and a standard deviation in LAI of 1.0. Then δ equals 0.4 and the 95^{th} percentile confidence interval for the ESU LAI is then [2.4,5.6].

3.2.5 Upscaling of Reference LAI Estimates

Accuracy assessment requires spatial matching of reference LAI and satellite LAI product estimates. While ESU LAI estimates can be used directly as reference LAI estimates without further spatial scaling, there are limitations to the approach. Firstly, satellite product mapping units can have length scales on the order of 1km while current in situ measurement methods are based on statistical models of vegetation density that may not hold as ESU size grows (Nilson 1971, Nilson 1999). Secondly, while ESUs with length scales on the order of 1km have been surveyed in landscapes with relatively constant vegetation patterns (Privette *et al.* 2002) the effort required limits the number of ESUs and hence both the precision and accuracy of comparison statistics. On a practical basis the current in situ LAI measurement networks use ESUs on the order of 1ha in area or smaller e.g. (Law *et al.* 2008).

Spatial scaling of ESU measurements can reduce uncertainties and increase the spatial extent of the reference LAI map. Spatial scaling methods generally fall into one of two categories: 1) data driven / structural and 2) functional approaches. Data driven or structural methods rely on latent variables (systematic unmeasured variables) to develop relationships between LAI and measured predictor variables. The simplest example being structural linear regression where there are measurement errors in both LAI and predictors. In this case latent variables are the underlying noise free predictors (Cheng *et al.* 1999). More complex data driven approaches include stepwise regression, mixtures of Gaussian models, neural networks and regression tree approaches that allow incorporation of multiple latent variables (Duda *et al.* 2000, Hastie *et al.* 2011).

Studies have investigated the use of spatial statistics (e.g. location) during upscaling through a family of kriging techniques (Garrigues *et al.* 2006, Martinez *et al.* 2010). The studies indicate that adequate performance with kriging methods requires a spatial covariate such as the normalised difference vegetation index (NDVI). In this case the additional gain from including spatial statistics may be small in comparison to simply imposing a landcover-based stratification. Nevertheless, it is good practice to include spatial statistics within structural scaling where there is a priori basis for their application (e.g. smoothly varying LAI patterns).

Functional upscaling involves use of a functional equation to relate predictor variables to LAI without the use of latent variables (Cheng *et al.* 1999). For example, a linear equation could be used to relate LAI to a spectral vegetation index. It is good practice to use functions that describe causal relationships based on physical principles. For example, simplification of the radiative transfer equation leads to a logarithmic functional relationship between LAI and the NDVI (Baret *et al.* 1991). A more complex functional approach is to perform non-linear inversion of a radiative transfer model whose parameters have been constrained using observations over ESUs (Verger *et al.* 2011).

Example 5: Structural Upscaling

The VALERI network (Baret 2012) uses a moderate (25 to 50) number of ESUs located within a 3km x 3km region. ESUs are located using a stratified random sample designed to match the cumulative distribution function of a surrogate for LAI (e.g. a vegetation index) over each dominant land cover class (e.g. forest, crop, grass). A robust multivariate structural regression is used to produce a transfer function relating multi-spectral measurements over the region with ESU LAI. The transfer function is only applied within the 3km x 3km region. This approach is good practice considering the high ESU sample density and the availability of prior information during sample allocation.

A typical VALERI site corresponding to an agricultural region in France is shown in Figure 5. While the spatial distribution of samples is relatively good, the comparison of the cumulative distribution of ESU NDVI to other random sampling trials indicates that very low LAI conditions are not sampled sufficiently. Normally the observed deviation would not be acceptable for a reference map. However, in this case, these areas correspond to bare fields that were assigned an LAI of 0.

The selected transfer function corresponded to a linear regression of LAI versus NDVI. As Figure 6 indicates, the transfer function provides relatively uniform distribution of residuals with a few outliers. As a result of the univariate regression most of the region falls within the spectral convex hull – the exception being the unsampled bare fields. The final LAI map indicates substantial spatial variability in LAI within the 3km x 3km domain that would otherwise make direct comparison between a single ESU or a few ESUs and a medium or coarse resolution satellite product difficult.

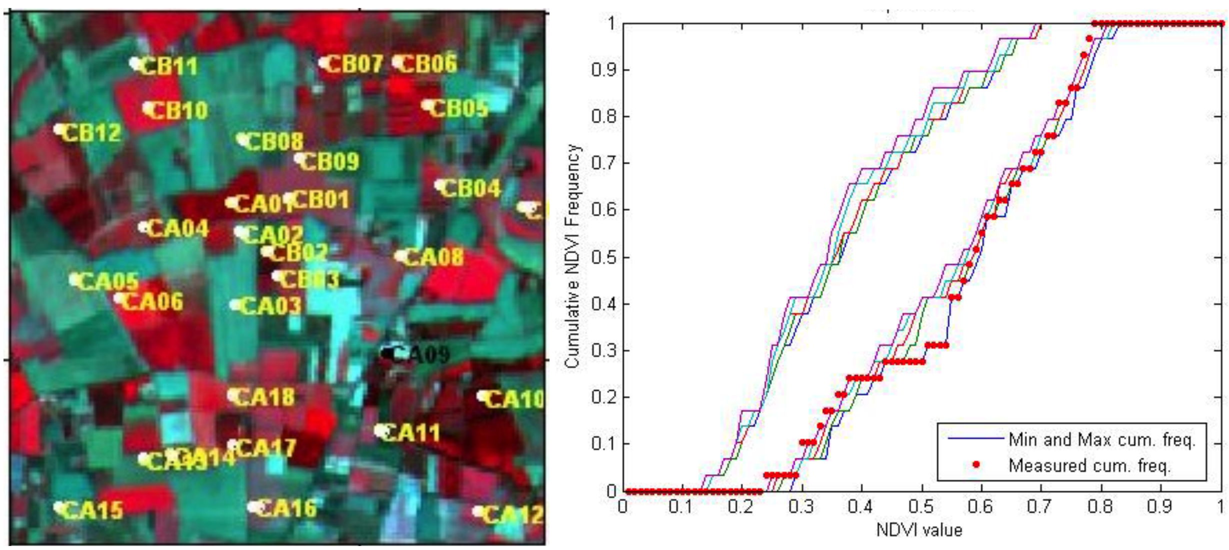


Figure 5: VALERI ESU sampling over a 3km x 3km agricultural region. Panel on left indicates location of samples over a colour composite satellite image. Panel on right compares the cumulative frequency distribution of sampled NDVI at ESUs (red dots) versus extreme ranges based on Monte-Carlo ESU sampling with the same sample size

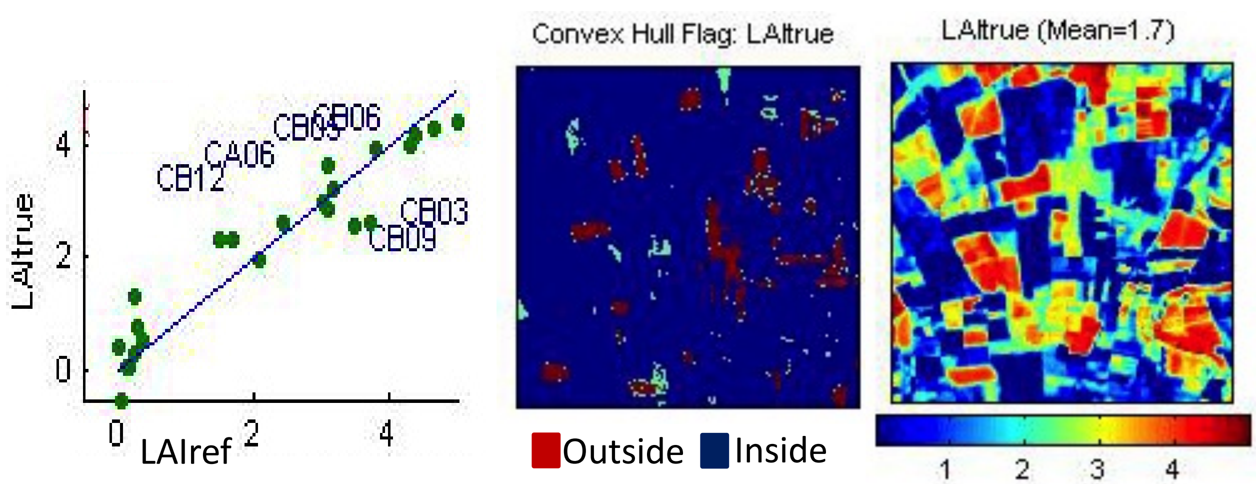


Figure 6: Outputs of VALERI reference LAI mapping process. Left panel shows scatter plot of predicted versus actual LAI based on robust linear regression. Outliers are indicated as alphanumeric symbols. Centre panel shows mask of areas within (blue) and outside (red)

spectral convex hull of ESU data. Right panel shows final reference LAI map.

Example 6: Functional upscaling

The CCRS approach for production of reference LAI maps relies on functional relationships between a spectral vegetation index and LAI for specific groups of land cover classes in a restricted geographic region (Fernandes *et al.* 2003). Figure 7 shows a region consisting of mixed boreal forest and a large open pit mine over which a reference LAI map is required. An initial field survey indicated either forest or shrub cover with moss or herbaceous understory and a maximum LAI of 4.0. Under such conditions separate linear functional relationships are expected between LAI and the reduced simple ratio (RSR) for needle leaf and broadleaf sites (Brown *et al.* 2000, Chen *et al.* 2002). Accessible regions for field surveys were identified in the north, centre and south regions of the study area to provide a spatial convex hull large enough to encompass the mining activity. For each region, ESUs were uniformly sampled with respect to RSR based on a previous satellite image. A total of 200 ESUs were sampled.

Separate transfer functions were developed using the Thiel-Sen functional regression (Fernandes *et al.* 2005) for broadleaf and needle leaf cover classes and applied to RSR data over the region using a land cover map. The systematic and random accuracy components were modelled as described in Section 3. Systematic error was generally below 0.5 units except in areas of high density broadleaf where the transfer function confidence interval was large due to a clustering of sampled ESUs around higher LAI values.

One could have adopted either more complex functional forms or additional input features. The decision to employ a univariate linear prediction was based on the following criteria.

Are the assumptions of the transfer function satisfied?
Figure 8 shows the transfer function used to relate LAI to RSR for needle leaf land cover using a robust Thiel-Sen regression (median absolute deviation 0.71; R^2 = 0.61). The regression assumption related to linearity was confirmed by visual assessment of residuals and by the fact that the 95th percentile confidence interval of the slope was non-zero.

Can the transfer function be improved by adding more features?
A multivariate linear regression (median absolute deviation 0.65; R^2 = 0.64) was also used to establish a transfer function of LAI as a function of two vegetation indices (simple ratio (SR) and infrared simple ratio (ISR)) that are also known to be linearly related to LAI over needle leaf forests for low to moderate levels of LAI (i.e. LAI up to 4.0). The multivariate regression provided a small decrease in residuals in comparison to the univariate regression but as figure 8 indicates, the prediction confidence interval doubled. The prediction confidence interval represents a spatially persistent accuracy error that will not cancel out during aggregation. As a result, the accuracy error of the aggregated reference LAI map based on both the ISR and SR features will be twice the magnitude of the map based only the ISR predictor.

Does adding more features substantially reduce the area within the convex hull of features?
As figure 9 indicates, the area falling within the convex hull of the SR and ISR features sampled by the ESUs is ~75% than the area when using only the RSR. More importantly, the reduced areas are spatially disjoint so that the area over which the reference map can be spatially aggregated using simple (e.g. rectangular) mapping units when using both SR and ISR will be substantially less than 75% of the area when using the RSR only.

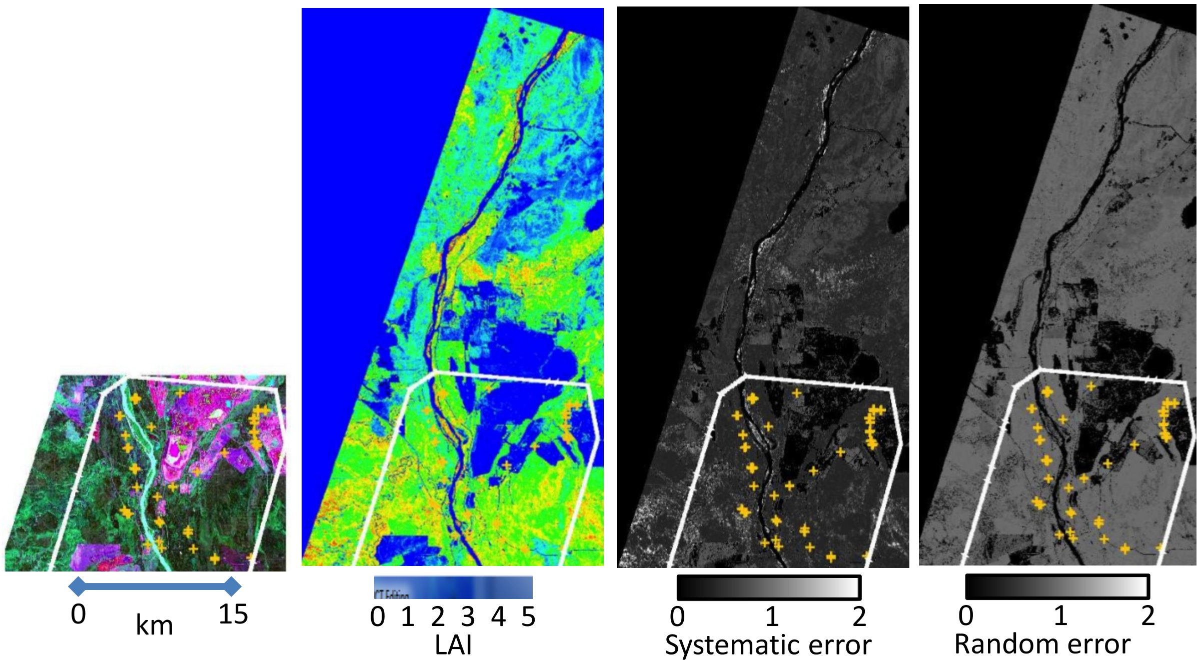


Figure 7: CCRS approach for reference LAI map production over a Boreal forest region with an open-pit mine (purple areas). ESUs are indicated as yellow crosses. The white outline corresponds to the spatial convex hull containing ESUs dilated by 1km. ESUs outside the convex hull did not correspond to land cover used for transfer functions (needle leaf or broadleaf forests).

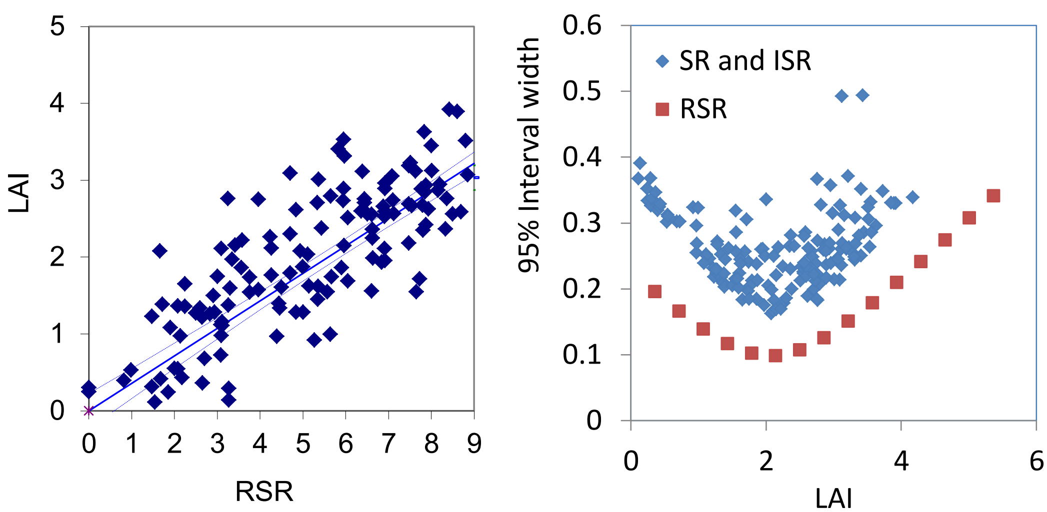


Figure 8: The left panel shows the transfer function for the needle leaf land cover class developed using a robust Thiel-Sen univariate linear regression (solid line) with its 95th percentile prediction confidence interval. The right panel compares the size of the prediction confidence interval of the univariate regression shown in the left panel with a transfer function based on multivariate linear regression using two vegetation indices (SR and ISR).

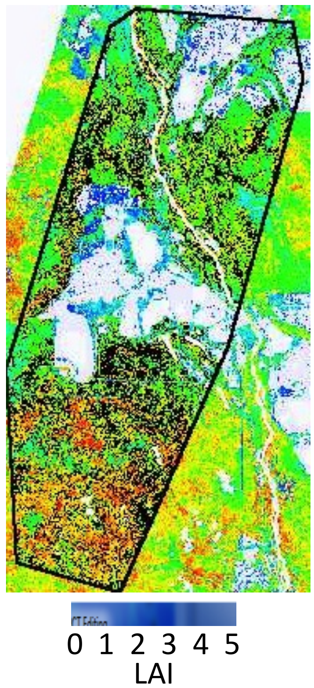


Figure 9: Convex hull of reference map area using only RSR (black border) together with areas falling outside the convex hull when using both SR and ISR mapped as black. White areas are designated as LAI 0 based directly on land cover.

3.2.6 Sample Size

Sample size, n, is a primary concern with reference ESU sampling given the expense of manual LAI measurements. Increasing sample size will generally increase the precision of the upscaling model and hence decrease the size of the confidence interval of predictions of reference LAI estimates. Smaller prediction confidence intervals lead in turn to smaller spatially persistent errors in the reference LAI map.

For a univariate linear functional model the confidence interval will decrease in proportion to $1/\sqrt{n}$ suggesting diminishing reductions in accuracy error after n~100. The same argument holds when using samples to validate reference maps produced using inversion of physically based radiative transfer models since their output LAI is assumed related to ESU in situ LAI using a 1:1 function.

A functional model with m predictor variables would require m^n samples to achieve the same precision (Good *et al.* 2009). This implies that ~1024 samples would be required for a 2 predictor linear model to achieve the same precision as a univariate model based on 10 samples. To our knowledge there has been no measurement campaign with over 1000 local ESUs. This strongly suggests that it is good practice that models with a minimum number of predictor variables be adopted unless their accuracy is significantly worse (e.g. the predictions of the univariate model exceed the 95^{th} percentile confidence intervals of prediction of the multivariate model). The choice of variables to use and the need for accuracy in their measurement may be determined from radiative transfer theory aided by site expertise.

For linear structural models where the transformation from predictor to latent variables is specified a priori the prediction confidence interval will have similar behaviour as for functional models with m now corresponding to the number of latent variables (Westland 2010). As such it is good practice that no more than one latent variable be adopted unless the model accuracy is significantly worse than with increased latent variables. However, some data driven approaches attempt to define multiple latent variables using parameters calibrated from observations (e.g. neural networks, regression trees, radial basis function methods). The calibration process will then require additional samples to provide the same prediction confidence intervals as the case where the parameters are defined a priori. For example, with partial least squares structural linear regression the additional sample size is proportional to the square of the ratio of predictor to latent variables and is bounded below by n = 700 (Westland 2010). To our knowledge there has been no measurement campaign with over 700 local ESUs. As such it is good practice not to use data driven approaches that require calibration of latent variables to define transfer functions (e.g. neural networks) unless these approaches provide significantly greater prediction accuracy than models with fixed latent variables (e.g. multivariate regression).

3.2.7 Sampling Design

Most reference networks are based on ESUs located for purposes other than LAI validation, so that ESU sampling is not designed optimally. However, it is worthwhile to identify preferred sampling designs in case the design is flexible or the network owners are prepared to change their design but also to identify the appropriate scaling approach for the available sampling.

Sampling design should satisfy the assumptions of the transfer function used to produce reference maps. All transfer functions require two fundamental sampling requirements to ensure they are unbiased over areas other than the sampled ESUs:

1. All population units have a non-zero chance of selection
2. Selecting one unit does not eliminate the possibility of selecting another unit

These requirements can be met by ensuring some level of randomness in the location of samples and ensuring sufficient sample size to avoid a trade-off in sample allocation such that a specific land cover within the reference area is completely ignored. In this sense it is good practice to stratify sampling by land cover and if possible LAI to ensure sufficient samples are present before the transfer function is calibrated. Moreover, the ESU locations within each stratum should involve sufficient randomness so that any LAI level in the stratum, but not necessarily any location, can be observed.

In some cases, sampling design can be further optimised to increase the precision of the transfer function based estimates. The simplest case is represented by the allocation of samples across different strata to minimise the variance of strata means. (Widlowski *et al.* 2007) developed an optimal allocation model to determine the number of samples n_h for stratum h:

$$n_h = \frac{N_h \sigma_h / \sqrt{c_h}}{\sum_{k=1}^{L} N_k \sigma_k / \sqrt{c_k}} \quad h = 1, \ldots, L \tag{7}$$

Where, N_h is the number of sampling units in stratum h, σ_h is the standard deviation within stratum h, c_h is the unit cost within stratum h, and L is the total number of strata. This approach requires prior specification of the cost of sampling in each stratum and the variance of LAI within the stratum. In the absence of this information this approach defaults to sampling proportional to the stratum area. Extending this approach to ordinary least square (OLS) structural regression leads to mean-sampling where the sampling is allocated proportional to the relative frequency of some indicator of the LAI distribution. Since a linear model is assumed between auxiliary variables and LAI the relative frequency can be estimated using one selected auxiliary variable.

Functional approaches assume that there is an underlying function relating regressors and LAI. The function may be simple (a linear equation) or non-trivial (a radiative transfer model). In either case a uniform sampling design is required to ensure a consistent or unbiased transfer function (Cheng *et al.* 1999). This implies that it is good practice to allocate samples to strata sampled uniformly over the range of LAI and land cover conditions rather than proportional to the area covered by each stratum. From a practical point of view uniform sampling will be impacted by the cost of samples. So a two stage approach may be required beginning with a modest number of uniform samples to meet the requirements of the functional regression, followed by additional replicates allocated based on cost (e.g. in accessible clusters) to increase the precision of the transfer function estimate (Cheng *et al.* 1999).

Cluster sampling can be a cost effective strategy for both functional and structural models as long as the cluster centres are located according to the optimal allocation rule for the model. In some instances clusters may be sufficiently dense to allow for a direct scaling of reference LAI based on their expected value (mean or median depending on the potential for outliers) that can serve as a local higher accuracy reference LAI value in addition to the larger area reference map.

To ensure that the accuracy of the transfer function applies during prediction it is good practice to restrict the prediction to the spatial, temporal and thematic convex hull of the calibration data. The spatial convex hull restricts the interpolation of reference map estimates to regions that should have similar climate and species conditions. The temporal convex hull restricts interpolation to the growing season i.e. conditions similar to the sampled ESU LAI. LAI reference maps can also be applied to anniversary intervals where there is prior information justifying the stability of LAI between different years for that interval. The thematic convex hull restricts interpolation to the convex hull of the data with respect to the joint distribution of auxiliary variables (Weiss *et al.* 2004, Martinez *et al.* 2010). This could correspond to the multispectral convex hull of the training data when multispectral images are used for scaling. It can also include consideration of nominal thematic partitions such as land cover. It is good practice to map the convex hull of the training data using the same auxiliary variable layers required for the reference map.

There is a fundamental tradeoff between the number of auxiliary variables and the area of the reference map falling within the training data convex hull due to the Hughes phenomenon (Hughes 1968). The Hughes phenomenon suggests that the proportion of the reference map falling within the convex hull will decrease as the number of independent auxiliary variables increases (Schowengerdt 1997). Global LAI validation implies the need to maximise the reference map sample size. As such it is good practice to use transfer functions that provide the largest mapped area within the training data convex hull as long as the reference map accuracy is sufficient to produce useful validation statistics. The exact tradeoff between accuracy and sample size is data dependent but can be assessed on a case by case basis. It is good practice to use the transfer function corresponding to the larger convex hull in the event that two transfer functions have similar accuracy.

3.2.8 Reference Map Accuracy

The reference map accuracy (technically this is the total measurement error, see section 5.2.1) corresponds to the size of the prediction error of LAI values in each mapping unit. The accuracy error can be reported as the median absolute deviation between predicted and actual LAI. However, since this deviation will vary with LAI it is good practice to report the accuracy error as a function of LAI (e.g. in ranges of 1.0 LAI unit). Accuracy error includes both spatially random and systematic components. Ideally one could employ a spatial analysis of residuals to quantify both components. ESU sampling density is usually insufficient to quantify the spatial pattern or residuals. When linear transfer functions are employed, the prediction confidence interval can be used to quantify the systematic component at a given LAI level. In this case, it is good practice to report both the systematic component of accuracy and the remainder after removing this component from the total accuracy.

3.2.9 Reporting of Statistics

Both data driven and functional upscaling approaches should report prediction accuracy statistics on a reference LAI mapping unit basis. It is good practice to report the 50^{th} percentile and 95^{th} percentile absolute residual as well as to plot residuals versus LAI for each stratum. Where possible, the confidence interval for the accuracy should be quantified using either resampling cross-validation or using the confidence interval of the linear prediction. Precision of the reference estimates is not a major issue since current validation protocols do not include resampling from the map. However, the spatial pattern of residuals is important since random spatial residuals would tend to cancel when upscaling to mapping units used for product validation. Generally speaking ESU sampling designs for most field experiments do not allow for detailed analysis of the spatial pattern of residuals (the exception being the BIGFOOT design reported in (Cohen *et al.* 2006)). However, to facilitate estimation of this bias during product validation, a table of residuals together with location and land cover condition should be included.

4 GENERAL STRATEGY FOR VALIDATION OF GLOBAL LAI PRODUCTS

4.1. Current Products

Global and regional LAI products are commonly generated from moderate (250m-500m) to coarse (>1km) resolution visible and near infrared (300nm – 2500nm) passive optical satellite imagery (Garrigues *et al.* 2008a, Camacho *et al.* 2011, Fang *et al.* 2013). More recently systematic production of medium resolution (10m – 100m) LAI products has been proposed (ESA 2007, Ganguly *et al.* 2012); see http://lpvs.gsfc.nasa.gov/producers2.php?topic=LAI.

Current systematic global LAI products require as input either bi-directional surface reflectance (BRF), directional-hemispherical surface reflectance (DHR) (black-sky albedo) or bi-hemispherical reflectance (BHR). Some LAI algorithms are specific to particular nominal acquisition geometry and therefore require a further angular normalisation process. All of the products require screening for contaminated pixels due to clouds, cloud shadows and detector failures. Often, this screening is either performed entirely, or at least refined, using multi-temporal composites based on spectral criteria. Such compositing will bias product retrievals to snow-free, peak vegetation density conditions in the composite window. Even with cloud screening and temporal compositing, most current multi-year products include a temporally smoothed version based on removal of outlier LAI values and temporal interpolation. Rather than discuss the hypothetical merits of each product, this document focuses on identifying how noise in the input data, geometric characteristics, and temporal aspects of products impacts LAI validation.

4.1.1 Uncertainties Related to Input Data

In this section sources of uncertainties in products resulting from input data are discussed. The majority of products rely on atmospherically corrected bi-directional reflectances – based on instantaneous multi-angular sampling and daily or multiple-day compositing of acquisitions with varying illumination-view geometry. It is relatively straightforward to generate error models for LAI retrievals as a function of input reflectance data uncertainties (Yang *et al.* 2006, Fang *et al.* 2013). For example, figures 10 and 11 show the spatial and temporal pattern of product retrieval uncertainties reported by producers for three different global products. However, the range of uncertainties is not entirely realistic both within a product (e.g. zero error in northern regions when snow is present) and between products (e.g. the JRC-TIP uncertainties can be much larger than other products over some land cover types due to differences in specification of uncertainty terms). It is likely that these uncertainties are often not perfectly specified, especially over varying land cover, atmosphere and acquisition geometry conditions. In such cases it would be useful to define a suite of canonical uncertainty scenarios. Table 5 offers a candidate range of parameters over which sensitivity analysis should be conducted.

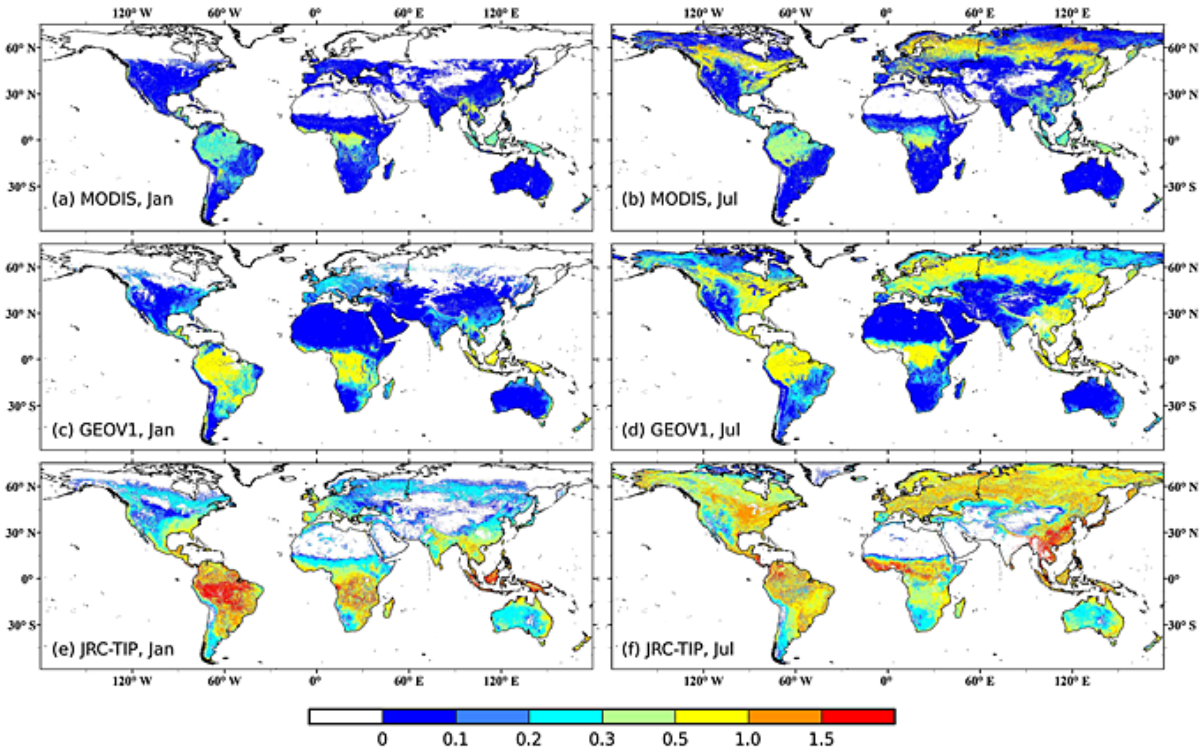


Figure 10: Average uncertainty for three global LAI products between 2003 and 2010 for January (left) and July (right) as stated by the producer (Fang *et al.* 2013).

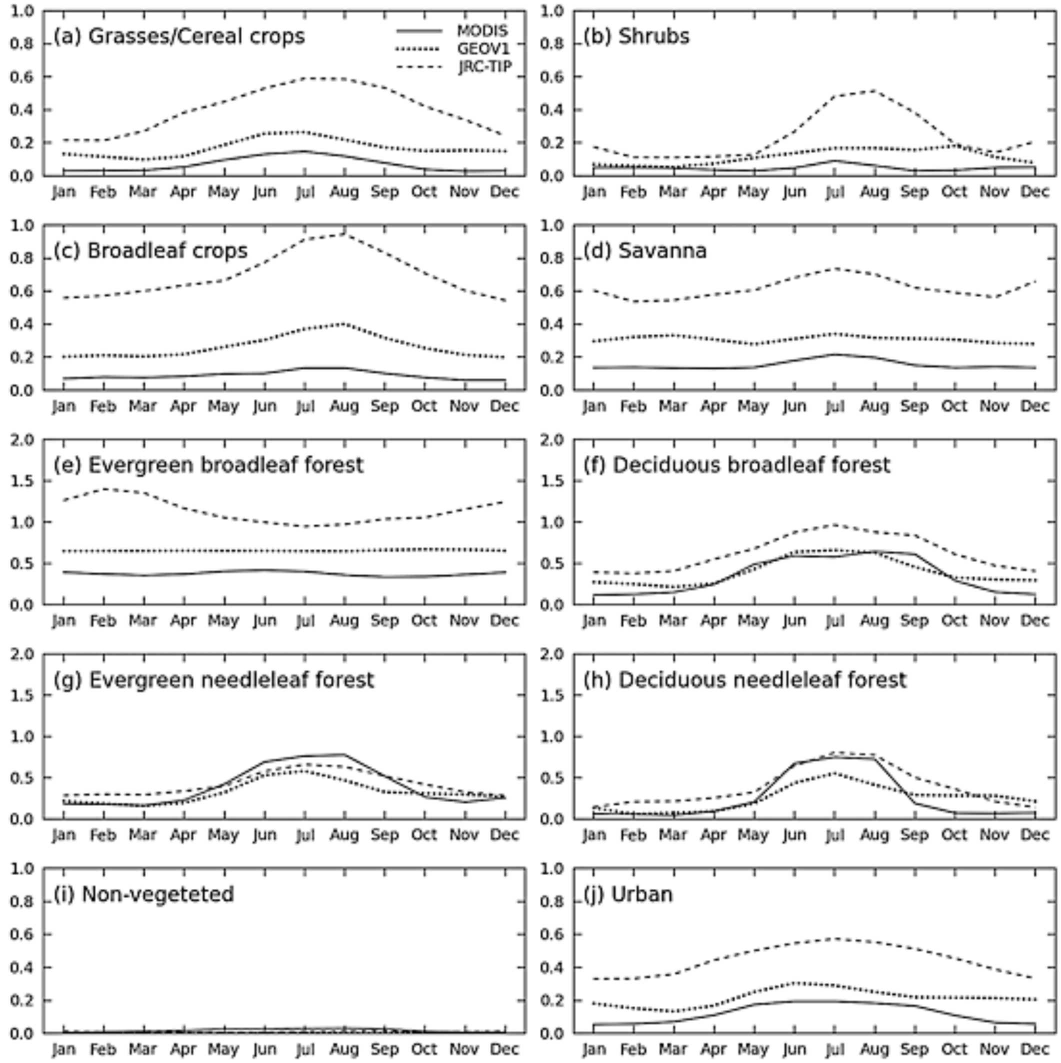


Figure 11: Climatologies of producer uncertainties for global LAI products between 2003 and 2010 over major global biomes. The y-axis is the uncertainty in LAI units. From (Fang *et al.* 2013).

Table 5: Recommended parameters with suggested ranges for sensitivity analyses for LAI products.

Quantity	Levels
LAI	0 through 10
Land Cover	Evergreen needle leaf forest, evergreen broadleaf forest, deciduous broadleaf forest savannah and shrubs and barren regions, herbaceous crop, forbes, wetland or regions with >25% water cover sub pixel
Radiometric calibration error	Sensor specific but within +/- 1.0 RMSE (root mean square error)
Geometry	0,35,70 degrees solar zenith angle for nominal and ±1 sigma view zenith angle in principal and perpendicular plane
Aerosol optical depth at 550nm	Uncertainty of 0.1 unit around a nominal level of 0.1 and a hazy level of 0.5
Canopy condition	±25% and ±50% uncertainty in specific chlorophyll, dry matter and moisture content
Understory reflectance	Sensitivity of LAI to change in understory from sand to clay to snow using standard reference spectra
Geolocation	Sensitivity to terrain slope of 10 degrees and 20 degrees

4.1.1.1 Sensor Noise

Modern optical imagers have sufficient quantisation and signal to noise ratio that sensor noise is far less of an issue than clear sky uncertainty. However, for very dark targets (i.e. dark soil, dense vegetation) and at low solar zenith angles, sensor noise effects should be quantified by sensitivity analysis of LAI algorithms applied to the scenarios given in table 4 including expected levels of sensor noise.

4.1.1.2 Clear Sky Uncertainty

In water vapour windows within the electromagnetic spectrum corresponding to spectral measurements used in most LAI algorithms, the clear sky top of atmosphere (TOA) radiance in visible and near-infrared wavelengths is impacted by: surface reflectance and its anisotropy (BRDF); the reflectance of adjacent surfaces; atmospheric transmission; and path radiance. Over flat surfaces and at moderate resolution (250m – 1km), uncertainties due to unaccounted acquisition geometry and transmission effects are typically smaller than the path radiance, especially in the visible and red-edge regions of the spectrum. Path radiance is driven primarily by aerosol properties suggesting that the impact of error in the specification of aerosol type and optical thickness must be quantified.

4.1.1.3 BRDF Modelling Uncertainty

Many current LAI algorithms utilise HRDF or BRF with observed sun and view geometries as input data and therefore, do not rely on explicit BRDF modelling (examples include MODIS, MISR and GEOLAND2 LAI products). For algorithms utilising nadir BRDF-adjusted reflectance (NBAR) and intrinsic albedo quantities (e.g. white-sky and black-sky albedo), the effect of BRDF model selection (e.g. linear, empirical, analytical) should be quantified across a range of surface conditions (e.g. flat vs. sloped surfaces) and dominant vegetation cover types. This can be performed using current global BRDF datasets (e.g. MODIS, MISR, and POLDER instruments) using ray-tracing simulations and by algorithm sensitivity analysis.

4.1.1.4 Canopy and Understory Modelling Uncertainty

There is currently an ongoing process (Widlowski *et al.* 2007) to radiative transfer models used for LAI retrievals. However, the process is still developing a framework for evaluating model inversion and for testing operational model inversion code. Prior assumptions regarding canopy leaf conditions and understory reflectance often have a large impact on LAI estimation and need to be quantified for their impact on LAI algorithm performance. In cases where a model that includes leaf and understory parameters is used, an algorithm sensitivity analysis should be conducted as well as comparison of retrievals over areas with rapidly changing understory reflectance (snow cover transitions) or leaf reflectance that can be qualitatively detected in imagery.

4.1.2 Geometric Considerations

Current LAI products are gridded in map projection systems rather than having each mapping unit associated explicitly with a geolocation field. Depending on the projection chosen, each gridded mapping unit (pixel) may have varying shape or area. For example, the Plate Carree projection (e.g. GLOBCARBON product, (Plummer *et al.* 2006) implies pixel width decreases with increasing latitude. In contrast, the Integerized Sinusoid Projection (e.g. MODIS product) will produce duplicate pixels with elongated parallelogram shapes when represented in Plate Caree projection at high latitudes. These considerations are important for both product application and validation and should be included in any spatial accuracy assessment.

LAI products are based on satellite measurements whose effective projected instantaneous field of view (EPIFOV), in general, will not exactly match the mapping unit. This occurs for three reasons listed below:

1. The change (usually growth) in pixel size further from nadir scan angles, often up to five times the nadir value for wide swath whiskbroom sensors.
2. Terrain effects change the shape, nominal location and to a lesser extent size of the ground projected instantaneous field of view (PIFOV). Certain processing chains (such as the MODIS MODAPS system) apply orthorectification to provide

a precise nominal location for all terrain. However, the majority of sensor processing chains do not include orthorectification by default. In any event no current processing chain accounts for the variable shape of the PIFOV.

3. Most LAI products are derived from reflectance measurements resampled to a final projection system and geoid. The resampling, often done with a bilinear interpolation or a cubic convolution, induces a low pass filter effect that tends to reduce information at high spatial frequencies and effectively decreases the scale of the information. All of these effects are a concern for the producer of LAI products. However, when one is concerned with diagnosing the benefit of different algorithms with the same input satellite reflectance data, they should be included in the final error analysis.

4.1.3 Temporal Considerations

GCOS specifies a target of bi-weekly LAI values derived from daily products. As such, ideal validation should include evaluation of differences at daily and bi-weekly time steps. However, current products range in reporting intervals from 4 days (Shabanov *et al.* 2005) to monthly (ESA 2007). Inter-comparisons can usually be performed by temporal aggregation of fine resolution products, but reference datasets correspond either to the day of the in situ data or the day of the auxiliary imagery used to scale up the in situ data. This represents a potential temporal mismatch that should not be attributed to an error for a product under assessment. LAI products should be inter-compared on a standard temporal aggregation interval and direct validation should be performed with nearest temporal neighbour interpolation where required. Each direct validation reference site should include the temporal variability expected due to temporal mismatch. This variability can be estimated using multi-temporal in situ measurements or by evaluating time-series of ensembles of LAI products.

4.1.4 LAI Product Definitions

Differences in the definitions of quantities represented in the currently available LAI products have led to substantial variability across performance assessments. In some cases, products and reference data both used LAIe leading to potentially optimistic performance statistics since clumping does not have to be included. In other cases products that retrieve LAI for both overstory and understory, as per the definition, were compared to reference measurements only from the overstory and hence tended to show overestimation (e.g. CCRS VGT total LAI versus MODIS collection 4 overstory LAI compared in (Garrigues *et al.* 2008a)). It is possible to qualify and discuss the comparison of LAI products and reference data that have different LAI definitions by carefully modelling these differences. However, it is clear from the perspective of GCOS that acceptable products should be compared to reference LAI datasets corresponding to the total LAI as defined by GCOS. Unfortunately, there are limited reference datasets available that actually report LAI – many correspond to an PAIe and others, even when corrected for clumping, are PAI due to the optical instruments used. It is good practice in this regard to specify clearly what the product definition is and conduct inter-comparisons taking into account any differences.

4.2 Status of Current Validation Capacity

Validation of global LAI products has progressed from producer driven studies to CEOS sponsored efforts to ongoing validation of operational products (GEOLAND, (Camacho *et al.* 2012); JRC-TIP, (Pinty *et al.* 2011)). Initially, producers such as the NASA MODIS Science team, CCRS Environmental Monitoring Section, GLOBCARBON (ESA Data User Element, DUE) and the CYCLOPES teams conducted product specific validation. Through the efforts of NASA/ORNL, ESA/INRA and CEOS, community accessible reference LAI data sets have been made available (http://lpvs.gsfc.nasa.gov/lai_intercomp.php). At the same time, a range of approaches for performing inter-comparison and direct validation have been developed (Morisette *et al.* 2006, Garrigues *et al.* 2008a, Camacho *et al.* 2010, Fang *et al.* 2013). From these approaches arise three fundamental components for a validation protocol: 1) direct validation over upscaled in situ reference datasets; 2) inter-comparison of products over a representative global sample; and 3) statistics related to the temporal completeness of LAI products.

4.3 Validation Requirements

A general validation strategy should be capable of testing products for compliance with GCOS requirements. A distinction is made between the strategy, corresponding to a sampling design, a definition of key reference datasets and inter-comparison methods, versus the data required to allow this strategy to test if products meet either threshold or goal requirements.

The validation strategy has five major criterion detailed in the following sub-sections:

1. Direct validation on a global basis representative of seasonal conditions and estimation of accuracy in LAI units.
2. Quantify the representative LAI accuracy estimate over areas or time periods without reference datasets.
3. Quantify the precision of LAI estimates over space and time on a globally representative basis.
4. Quantify the long term (inter-annual) stability in LAI products.
5. Identify issues with algorithms or datasets that may cause biases in LAI retrievals.

4.3.1 Direct Validation on a Global Basis Representative of Seasonal Conditions and Estimation of Accuracy in LAI Units

Direct validation should be performed using available reference datasets traceable to in situ reference measurements accompanied by an associated assessment of their uncertainty. Estimates should be upscaled using one of two auxiliary variables: high resolution land cover maps or high resolution satellite/airborne measurements. The upscaling should take into account the possibility that the upscaled product contains regions where in situ data are not representative. One of three strategies should be used for filling in these regions: 1) heuristic estimates from look up tables (e.g. water has LAI 0); 2) estimates based on a biome specific radiometric relationship; and 3) estimates based on a maximum and minimum possible range of LAI. Gap filling of reference maps should be limited in spatial extent (e.g. only for small linear features such as roads, streams, pathways). In addition uncertainty in the auxiliary variable used to scale in-situ LAI should be accounted. Reference LAI maps should be produced using an approach that follows one of the upscaling protocols listed in Section 3.2.2 or is justified as being equivalent or superior to those protocols in terms of accuracy. The reference maps should be aggregated so as to minimize the impact of geolocation and binning uncertainty when performing comparisons with products. Paired values of spatially and temporally coincident product and reference values should be tabulated. The global subset of paired values should be compared using appropriate robust statistics and visualisation of residuals (Section 5.2).

4.3.2 Quantify the Representative LAI Accuracy Estimate Over Areas or Time Periods Without Reference Datasets

There are three issues with representativeness.

1. The precision of the accuracy estimate assuming the reference data are globally representative.
2. The spatial extent of the comparison.
3. The temporal domain that the comparison applies to.

The precision of the accuracy estimates themselves can be modelled using the confidence interval accuracy statistics. The spatial representativeness of accuracy statistics can be quantified in two stages. Firstly, the spatial variability of the product accuracy over subsets of a biome with different reference datasets can be evaluated. Secondly, recognising that reference data may be a biased global sampling, a diagnostic of spatial and temporal representativeness of the accuracy statistics is required. There are two diagnostics proposed. The first order approach is to include areas with the same land cover, LAI and seasonal conditions as representative regions. A more rigorous approach is to use areas that have similar agreement with a global, seasonally continuous ensemble reference as with the areas with reference measurements. Currently there are insufficient independently generated global LAI products to support such an ensemble and hence inter-comparison to regional LAI products with higher accuracy and consistency are suggested in Section 5.2.

4.3.3 Quantify the Intra-Annual Precision of LAI Estimates Over Space and Time on a Globally Representative Basis

LAI precision corresponds to variation in LAI estimates under constant in-situ conditions. Temporal precision will be evaluated by two approaches. For intra-annual precision, the deviation from linearity of midpoints of triplets as in (Camacho *et al.* 2012) is suggested. Spatial precision should ideally be quantified by computing the change in LAI over areas known to have temporally stable LAI patterns with high spatial variability. This information is not currently available so precision is currently estimated by statistics derived from the ensemble of regional correlations between successive product time slices. This represents a relative metric for comparing products until areas known to be temporally stable can be identified.

4.3.4 Quantify the Long Term (Inter-Annual) Stability in LAI Products

Current validation studies rely on a few (<10) sites with intra-annual measurements (e.g. (Garrigues *et al.* 2008a, Camacho *et al.* 2011, Ganguly *et al.* 2012). The majority of these sites are located at flux towers and ecological sites and with insufficient spatial sampling to provide high confidence in validation of current coarse resolution products. Until methods for continuous in-situ LAI survey are widely implements it is suggested that inter-annual stability be assessed by evaluating trends in product LAI over 3 x 3 pixel targets that are expected to have vegetation cover during the local growing season. Further study is required to identify such targets and to identify the appropriate trend metrics to quantify stability.

4.4 Challenges to Validation Strategy

4.4.1 Insufficient Reference Data

The limited availability of reference LAI data is fundamental, but not limiting to conducting global LAI validation. Spatial representativeness will grow with less expensive in situ survey methods and increasing use of LAI products in climate models. Temporal representativeness may be enhanced if phenological monitoring networks (such as the National Phenology Network (NPN) USA, Phenological Eyes Network (PEN) Japan and PlantWatch Canada) can be adapted to provide LAI information. Fortunately, there continues to be improvements in the representativeness of in situ LAI observations at flux towers e.g. European Initiative Integrated Carbon Observing System (ICOS) and US initiative National Environmental Observation Network (NEON).

4.4.2 Insufficient Products to Generate an Unbiased Ensemble

The small number of LAI products is a fundamental challenge for inter-comparisons since any ensemble of LAI products is likely to have biases due to both the small sample size and the fact that many products are related. This challenge could be lessened in the future by performing ensemble retrievals for products themselves or simply incorporating error statistics when computing the ensemble. However, at present we note that even the IPCC (Inter-governmental Panel on Climate Change) General Circulation Models (GCMs) were originally a handful of models sharing very similar designs. The size of the ensemble should be incorporated in inter-comparison statistics.

4.4.3 Thematic Differences in LAI Definitions

Progress has been made recently in harmonising the LAI definition between CEOS, GTOS and GCOS. However, satellite products still differ in their quantities and in general do not match this definition. The provision and ongoing refinement of transfer functions between different LAI quantities as well as dedicated efforts on product transparency will provide traceability for comparisons. See Section 2.1 for the definition agreed between CEOS, GCOS and GTOS.

4.5 Status of Current Validation Capacity

4.5.1 Data

Through the efforts of the CEOS LPV sub-group and its member Space Agencies, as well as a number of research groups, the global community now has open access to high spatial resolution (250m - 10km) global LAI datasets. However, these datasets have limitations. For example:

- They include variable quality control levels, usually with no uncertainty bounds.
- They are for limited time periods that do not all overlap.
- Their LAI definitions are not always consistent across biomes.
- They sometimes have limited valid retrievals under snow conditions.
- They do not share consistent land-water masks.

Nevertheless, (Garrigues *et al.* 2008a) demonstrated that a useful multi-year global inter-comparison was feasible at 10km resolution. This has been supported by further inter-comparisons by the GEOLAND2 project (Camacho *et al.* 2010, Camacho *et al.* 2011).

The ESA OLIVE system (On Line Validation Exercise, http://calvalportal.ceos.org/web/olive, (Weiss *et al.* Submitted)) has been designed for CEOS as a public system to allow producers of biophysical products to test their outputs against equivalents and against a set of validation data. OLIVE hosts extracts of current global LAI products over the sites shown in figure 1 (BELMANIP2 and DIRECT) (http://calvalportal.ceos.org/web/olive/site-description). The system implements a suite of comparisons including, where technically feasible, diagnostics recommended in this protocol.

The state of reference datasets is currently not as well developed as the satellite-derived LAI products. International flux tower and LTER (Long Term Ecological Research) networks quantify LAI, but usually over areas far smaller than $1km^2$. These data are archived within their networks but are not easily accessible to CEOS or LAI product producers. While GCOS has provided protocols for in situ LAI survey at local sites (Law *et al.* 2008), these methods may not be efficient for LAI validation. In contrast, regional and coordinated global networks (reviewed in (Morisette *et al.* 2006)) now provide consistently observed reference LAI fields at a small number (<100) of regions ranging from ~10km^2 to ~100km^2 (see Section 3.2). A concern with the regional networks is that they are not frequently revisited (for example, an annual visit to an evergreen forest is recommended to ensure the LAI field is constant over time).

There are currently diverse data sets that could potentially contribute to a global reference dataset. These fall under the following categories:

1. Local estimates of LAI or related quantities that cannot easily be scaled to 3 x 3 pixel regions.
2. Local estimates of LAI that have been scaled to at least 3 x 3 pixel regions.
3. Datasets that identify the absence of any vegetation (LAI = 0) or the presence of vegetation above a certain threshold LAI.

Under the first category fall data collected within the global FLUXNET and LTER networks, the regional FOREMON network, and coordinated regional or global scientific experiments. These larger communities have data access points and available databases. In contrast, many regional scientific experiments collect LAI data that is not catalogued in a common area and are therefore difficult to access for global validation.

Under the second category fall data collected for global and regional LAI validation. Most of these data are catalogued if not archived within the VALERI and ORNL databases although some agencies (NASA, CCRS, CSIRO, ESA) have funded projects that keep their own databases. Typically, these datasets correspond to ~30m resolution LAI surfaces produced by calibrating optical satellite imagery with in situ measurements over ESUs located using a stratified random sampling. In some cases (e.g. CCRS) the data are provided in a database with a relatively large (>100km^2) spatial coverage but with a mask indicating the region the data are deemed representative.

The third category includes data that could be used for LAI validation, but would first require a consensus that they are indeed accurate descriptions of vegetation conditions. The foremost would be available water surface coverage datasets both globally and at continental scale. While there may be some variability in coastlines one can expect that a reasonable spatial buffer could eliminate such uncertainty. Secondly, would be glaciers, as recorded by the World Glacier Monitoring Service through the Global Land Ice Monitoring from Space (GLIMS) activity. Thirdly would be deserts and exposed rock areas mapped in global and national land cover maps. These three datasets would require systematic screening for long-term changes but, on a global scale, should be probative of zero LAI conditions throughout the year. In addition, one could also take advantage of current maps of dense evergreen vegetation. In this case, a safe minimum LAI threshold could be established based on coincident local or image based estimates.

Data from the first two categories can, in principle, be combined during comparisons if care is taken to account for spatial mismatch between local data and global products and for the large sample size of many regional products. However, separate comparisons should be performed with the third category both because of the large, bi-modal, reference samples and of the likelihood some of the regions may have changed or been incorrectly mapped in the first place. We note here that the robust statistics proposed below implicitly address some of these effects.

4.5.2 Methods

Methods for LAI validation fall under two categories: production of in situ reference estimates and performance assessments. This section describes the major approaches that have been used for both of these activities.

4.5.2.1 In Situ Reference Estimates

Reference estimates require a sampling protocol (instrument settings, measuring sequence at a given point), a method for measuring LAI for each ESU and a method for upscaling ESUs to generate a regional LAI reference map.

Commonly, reference sites are located near infrastructure permitting easy access for field teams. This ad hoc approach was rehabilitated by CEOS by supplementing these sites to provide a globally representative sampling across land cover and peak LAI levels using existing experimental networks and the GLC2000 land cover map (BELMANIP) (Baret *et al.* 2006). This has recently been revised to be independent of ground experiment measurements as well as better represent the variability of vegetation types (using the GLOBCOVER 2009 land cover map) and climatological conditions at the Earth surface. Hence the network contains more homogenous sites that are appropriate for inter-comparison of products at 1km. The new version, BELMANIP2, is available from the CEOS Cal/Val Portal (http://calvalportal.ceos.org/web/olive/site-description). Additionally, in some areas (e.g. Canada, USA) biome or ecozone stratification is conducted for reference sites. However, these sites are infrequently visited. New reference sites mostly correspond to other funded projects requiring LAI information. Elsewhere, additional sites are being located at new flux towers and ecological monitoring sites, with little consideration for spatial representativeness and upscaling for satellite validation studies. It is expected that these reference sites and the BELMANIP framework will evolve as new sites that are appropriate for satellite validation exercises become available.

Sampling within reference sites has ranged from:

- Extremely dense spatial sampling that provides accuracy at the cost of efficiency (e.g. BIGFOOT).
- Sampling focusing on covering a flux tower or ecological mapping unit that is typically less than 1km^2 (e.g. FLUXNET).
- Stratified random sampling over a ~3km x 3km region (e.g. VALERI).

- Spatially replicated clusters of stratified random sampling together with samples along linear access routes (e.g. CCRS).

Spatial statistics have been exploited for scaling in situ ESU LAI over reference sites (Garrigues *et al.* 2006, Martinez *et al.* 2010). However, (Garrigues *et al.* 2006) report that that there is little benefit in exploiting spatial correlation structures not available from simple land cover segmentation. This finding supports the positioning of ESUs within relative homogenous regions on the order of 1ha in area that are then scaled using land cover or ancillary imagery although more sophisticated spatial statistics may be exploited in areas where samples are already allocated across continuously varying LAI patterns (Martinez *et al.* 2010).

Two general categories of upscaling approaches have been applied to generate reference LAI maps from ESU level LAI data (Section 3.2). Over regions without prior knowledge about relationships between scaling variables (such as reflectances) and LAI, structural approaches are required to define scaling transfer functions (e.g. VALERI). However, structural approaches are sensitive to measurement errors and typically less efficient than functional approaches that can assume *a priori* relationships (Fernandes *et al.* 2005). Although conceptually straightforward, there has only been recent work using calibrated radiative transfer models for reference LAI retrieval (Heiskanen *et al.* 2011, Verger *et al.* 2011, Claverie *et al.* 2013). This approach is promising in that it only relies on ESU information to define the range of model parameters and to validate the final maps.

Perhaps one of the most contentious issues in LAI validation is the accuracy and precision of ESU level LAI estimates. ESU estimates are challenging both because most must be non-destructive as well as account for woody material and spatial clumping. The GCOS Vegetation Survey protocol (Law *et al.* 2008) provides reasonable approaches for ESU LAI estimates based on destructive sampling for crops and forbs and optical sampling. These approaches are very time consuming and likely not feasible over large regions or for regular repetition. Moreover, they require use of two instruments (the LAI-2000 and TRAC) for PAIe estimates in forests and savannah. These approaches will not easily capture very low vegetation and it is not clear how accurate the TRAC clumping index estimate is when applied to LAI-2000 data for forests. These latter estimates also require correction for shoot clumping (Stenberg 1996) and woody area that can have large (>50%) uncertainties between and across species.

Many groups now make use of digital hemispherical photography (DHP) and digital camera photography (DCP) for ESU level LAI estimates. A number of studies indicated that such methods, while suitable for broadleaf canopy PAIe were limited by technical aspects of the cameras for short vegetation and dense forests (Zhang *et al.* 2005, Ryu *et al.* 2010). It is noteworthy that these studies often used non-professional grade instruments without calibration for geometric and radiometric responses. Work at INRA, CSIRO and CCRS suggest that professional grade cameras are unbiased and relatively precise (<20%) for broadleaf vegetation and can provide consistent PAIe estimates with some bias in clumping over forests with very elongated crowns (Weiss *et al.* 2004, Demarez *et al.* 2008, Pekin *et al.* 2009, Macfarlane 2011). Moreover, such cameras now have better resolution than passive direct beam instruments like the TRAC, especially when used in planar zoom mode rather than with fisheye lenses. Certain manufacturers also allow for correction of chromatic aberration and poor exposure leaving only focus and depth of field as issues for optimisation in situ.

Work within the VALERI project and in Finland indicates that 10-15 DHP images (or LAI-2000 measurements) are sufficient for PAIe in an ESU (Demarez *et al.* 2008, Majasalmi *et al.* 2012). Notwithstanding these developments, the issue of estimating woody area is still best addressed only over deciduous woody areas or by allometric equations. The censoring of woody matter in digital images may be appropriate for trees having distinct crowns, but is problematic in general since an unknown amount of foliage is also censored. CEOS has identified the need to establish systematic in situ monitoring activities appropriate for satellite validation across all vegetated regions under the BELMANIP framework (Baret *et al.* 2006). In future editions of the GCOS Satellite Supplement, this should be reinforced (see Actions T29, T3 and T4 within (GCOS-138 2010)).

4.5.2.2 Statistics Used for Performance Assessments

LAI validation studies have used common approaches for reporting product performance although the spatial and temporal extent of the sampling distribution used for comparisons has varied from local to global studies. Local studies have historically relied on exact spatial and temporal matches to reference LAI estimates. While useful for diagnosing specific issues with products, local studies do not serve to meet the GCOS reporting requirements or general user needs for performance over regional to global extents. For example, NASAs MODIS validation effort was based on a global biome stratification with regional sites optimised for access and the presence of local collaborators (Morisette *et al.* 2006). In contrast, regional and global validation studies now rely on *a priori* sampling designs for inter-comparisons. At a national scale, CCRS uses a stratified sampling approach based on selecting Landsat World Reference System frames that cover ecozones (one level more detailed than a biome) across Canada and maximises the match to within ecozone land cover distribution (Fernandes *et al.* 2003). The US National Environmental Observation Network (NEON) is also acquiring LAI estimates using a nested spatial sampling over each of 80 sites stratified by biome and climate zone across the US (NEON 2009). The BELMANIP2 sampling plan extends this approach to a global coverage.

With respect to direct comparisons, global LAI estimates are typically compared over co-located regions (3km x 3km or coarser) through scatter plots across global datasets, across biomes, or sometimes across a local region with extensive validation data. Statistics are reported to describe the bivariate distribution of reference and product LAI. These include the Pearson’s correlation coefficient, the mean relative and raw residuals (either as absolute residuals or squared residuals) and the bias from a 1:1 line of a linear fit (e.g. (Garrigues *et al.* 2008a)). The majority of studies surveyed do not test for normal or homeoscedastic uni-modal distributions of residuals. This should be conducted especially at global scale because it is likely residuals differ for different LAI levels that are themselves correlated with biome type or seasonal sampling. Claverie *et al.* (2013) specifically report residuals as a function of time using box-plots to address this concern.

Direct comparisons of temporal trends in LAI are extremely limited and often assume a small (<1km^2) site is representative of a larger pixel (Ryu *et al.* 2012). This assumption may be appropriate for large fields and homogenous forests but requires care with the in situ protocol and upscaling of reference data to ensure the uncertainty due to this assumption is quantified. For example, the CCRS protocol makes use of replicate ESUs within large farm fields and the work of the GEOLAND2 team relies on spatial statistics to account for heterogeneity in LAI patterns (Martinez *et al.* 2010). Ideally a time-series of reference LAI maps could be produced by applying the appropriate scaling transfer function to the time-series of satellite images. For example (Claverie *et al.* 2013) used a radiative transfer model inversion using neural networks as a transfer function applied to time-series of Formosat images over an agricultural region. Careful field measurements were used to quantify the accuracy of the reference maps over time. The combination of this approach with an automated in situ survey may facilitate future temporal accuracy studies.

Inter-comparisons are commonly conducted at a lowest common resolution or coarser to minimise differences due to spatial mismatch (Garrigues *et al.* 2008a). Time-series plots over selected sub-sampled regions, usually by land cover and biome type, are presented together with maps of inter-product differences on a monthly or seasonal basis. Scatter plots are sometimes presented, but we note that these are not commonly applied to the same sampling distribution used for direct validation.

Accuracy assessments typically include statistics describing conformity with assumed LAI patterns with an emphasis on temporal precision. These include distributions of LAI changes over dense evergreen forests and non-vegetated areas and computing deviations from local linearity in consecutive LAI retrievals over seasons. Studies evaluating spatial precision are uncommon, although (Fernandes *et al.* 2003) used Landsat Enhanced Thematic Mapper Plus (ETM+) and Thematic Mapper (TM) imagery with the same algorithm as their continental scale product to evaluate differences caused by using coarse resolution imagery. In principle, studies relying on high resolution imagery could also evaluate spatial precision assuming the global LAI algorithm could be applied to that image data set.

5 RECOMMENDED APPROACH FOR GLOBAL LAI PRODUCT VALIDATION

The goal of this section to provide guidelines for producing statistics related to the accuracy, precision and completeness of LAI products with global coverage.

Accuracy estimates require comparison of corresponding product and reference LAI values. The representativeness and the confidence interval of accuracy estimates are limited *a priori* by the current spatial and temporal coverage of reference LAI data. Previous validation studies have performed accuracy assessments using both pooled global reference datasets and comparisons to regional reference datasets (Garrigues *et al.* 2008a, Camacho *et al.* 2011, Camacho *et al.* 2012, Fang *et al.* 2013). Pooled datasets are problematic since: a) the sampled in situ sites cannot be assumed to be either representative of the global LAI distribution; b) LAI products are frequently based on biome specific algorithms that should then be validated at the biome level; and c) validation statistics from pooled global reference sites may be biased since differences between reference and products may be systematically related to land cover or climate conditions, while common accuracy statistics usually assume that differences arise from simple (e.g. bivariate normal, unimodal) distributions. In this sense it is good practice to quantify accuracy on a spatially stratified basis and to account for the reduced sample size by reporting the uncertainty of the accuracy statistics.

Estimates of precision can be derived using ensembles of LAI estimates from the same algorithm over the same surface condition (e.g. LAI, biome, climate zone). As such, precision can generally be estimated over global multi-year extents of a given product. The uncertainty in precision statistics is related chiefly to the assumption of similar surface conditions and to the temporal extent of the product.

Estimates of continuity can be derived directly from product metadata or quality flags, although care is required to account for differences in temporal sampling or data quality levels when inter-comparing continuity statistics.

LAI products often include uncertainty estimates generated based on theoretical or prior error models (Fang *et al.* 2013). Where available, these uncertainty estimates can be used to improve estimates of accuracy, but should not be confused with the validation statistics used here, that rely on comparisons of products and reference data.

5.1 Reference Data Sets

CEOS validation requires reference LAI measurements generated from methods independent of the products being validated. CEOS allows for both in situ and other suitable reference data. In this section we describe good practices for the selection of reference datasets, matching reference data sets to products and for producing and reporting accuracy statistics.

5.1.1 Reference Estimates Traceable to In situ Measurements

Ideally, a globally representative and continuously revised network of spatially extensive reference in situ sites should be available for global LAI validation. Until such a network is established, a practical validation strategy should make use of all available sources of reference data. As described in Section 3, three sources of reference estimates traceable to in situ measurements are available:

1. LAI measurements over individual ESUs.
2. Spatially extensive LAI reference maps based on data driven relationships calibrated using ESU LAI.
3. Spatially extensive LAI reference maps based on functional relationships calibrated using ESU LAI.

Good practice includes use of all three of these reference datasets as long as their thematic content is independent of the derived LAI products.

Information related to the performance (accuracy, precision, completeness) of reference datasets are useful both to ensure traceability to in situ measurements and for LAI validation statistics. Accuracy and completeness are most important since product validation often involves aggregating reference datasets.

Aggregation will reduce precision errors but accuracy errors that are not spatially random will remain. It is good practice to map the spatially random and systematic components of reference LAI error. Where reference errors are normally distributed the root mean square error (RMSE) or one standard deviation interval are appropriate error statistics for the reference map. However, reference map errors may include occasional outliers due to factors such as errors in land cover data used for scaling in situ LAI measurements. In such cases, the RMSE or one standard deviation interval may be pessimistic and a percentile interval is preferable. In this case the median absolute deviation and the 95^{th} percentile residual are recommended as estimates of typical and worst case errors in the reference map.

Completeness refers to the temporal and spatial extent over which the reference data can be traced to in situ measurements. Generally speaking reference measurements are conducted over a short period of time during which LAI is assumed to be constant. Nevertheless, to improve consistency of product evaluations the length of the survey interval should be documented and where possible the change in LAI during this interval included in the systematic accuracy error. Spatial completeness includes two aspects: 1) the spatial support of the measurements; and 2) the traceability of the method used to produce reference data to in situ measurements. For the first aspect, it is good practice to restrict reference data use to the spatial convex hull of measurements. The second aspect primarily applies when maps are produced using ancillary layers related to land cover or spectral imagery. In this case, it is good practice to limit the reference dataset to the land cover conditions sampled by these measurements in addition to addressing the first aspect. To facilitate reporting of completeness information it is good practice to provide a polygon layer indicating spatial regions and a calendar indicating data ranges where the data are complete in the sense described here. Examples 5 and 6 in section 3.2.5 indicate good practices for documenting the accuracy and spatial completeness of reference LAI maps.

5.1.2 Heuristic Reference Estimates

In many areas of the world, a suitable upper and lower bound on LAI estimates can be derived. For example, water bodies, deserts and glaciers have zero (or very low) LAI. In contrast, dense evergreen rainforests will always have some non-zero LAI as long as they remain forested. Table 4 indicates a range of in situ LAI available in global databases as a function of land cover class. In the absence of local measurements these ranges can be used to model the upper and lower bounds of reference LAI value. The uncertainty based on these ranges will be large but the net impact during validation may be small if the heuristics are limited to isolated cases (e.g. <1% of any validation mapping unit) or to cases where only bare areas are specified as zero LAI. The use of heuristics based on land cover will include an additional uncertainty due to land cover map errors. Good practices to model this uncertainty are not yet available, however, the uncertainty for a given land cover class and given land cover map should be reported. The classification confusion matrix should be included with this reference data to facilitate production of accuracy statistics once these guidelines are produced.

5.1.3 Co-location of LAI Estimates

Direct validation requires comparison of co-located reference and product LAI estimates. Differences observed in such comparisons will include both temporal uncertainty and spatial uncertainty when matching of reference and product LAI estimates. Part of these differences may be due to temporal or geolocation errors in product maps and should be included in accuracy assessments where possible. However, good practices are required to reduce the contribution of errors in reference map geolocation (geolocation uncertainty) or simply due to imperfect alignment of reference and product mapping unit (binning uncertainty).

Measurements within match-up units should ideally be independent of those in other units to simplify accuracy assessment. However, reducing geolocation error impacts on match-ups may result in large match-up mapping units and therefore fewer validation samples with less thematic detail (e.g. over mixed cover landscapes there will be fewer pure land cover conditions). As a compromise, it is good practice to minimise the spatial overlap of match-up mapping units and to then reduce the number of degrees of freedom on validation statistics proportional to the overlap.

Co-locations should be performed over a match-up mapping unit corresponding to a continuous spatial and temporal extent that encompasses at least one product estimate (e.g. a pixel) although it is good practice to use multiple product LAI values within a match-up mapping unit to reduce binning uncertainty. To facilitate reporting and replication of validation exercises, it is good practice that all reference and product LAI measurements / estimates in a match-up mapping unit should be extracted (instead of simply a summary statistic) together with their accuracy without further manipulation (aggregation, subsampling etc.) and stored in a database to facilitate independent verification of product performance. For consistency, it is good practice to adopt a local projection with low distortion in area across each reference region and to align product and reference maps along the North-South axes.

5.1.3.1 Geolocation Uncertainty

Geolocation uncertainty corresponds to imperfect specification of either the reference or product mapping unit boundaries. Causes of geolocation uncertainty include errors in locating mapping unit centroids, variation in the footprint of the sensor measurements and uncertainty due to resampling the measurements onto the Earth.

Spatial matching of product and reference mapping units depend on the size of the mapping units and their geolocation uncertainties. For both product and reference maps, the total spatial uncertainty should be quantified as the sum of the size of the mapping unit and the geolocation uncertainty. For irregular (polygon) reference units, the mapping units should be approximated as the minimum bounding rectangle oriented north-south. To minimise the impact of geolocation uncertainty on validation statistics it is good practice to perform co-locations over a mapping unit corresponding to the worst case geolocation uncertainty over all product and reference mapping units.

For simplicity it is good practice to use a north-south oriented rectangular moving window when performing spatial aggregation to match product and reference mapping units. The window should have dimensions of three times the length and width of the larger of the product or reference mapping units plus the 95th percentile worst case geolocation error. The exception being when comparisons are performed and individual ESUs fall within a product pixel. In this case the window size is the 95th percentile geolocation error otherwise binning error will become an issue otherwise due to the small spatial extent of typical ESUs.

5.1.3.2 Binning Uncertainty

In most cases, both the product and reference mapping units will not have exact spatial coincidence (e.g. different raster grid orientation or polygon reference maps and raster products) so some sort of binning is required to extract values over the mapping units used for comparisons. Considering that substantial resampling and interpolation may already have been performed to generate these maps, it difficult to optimise this binning for each possible product / reference data situation. Two common cases are considered here.

Firstly, the product or reference maps may not be complete over a match-up mapping unit. For example, comparing a small reference ESU with a larger product pixel containing it, or where the reference map has areas that are not represented by the transfer functions used to produce the map. In this case, it is good practice to use single pixel rather than 3 x 3 pixel mapping units for comparison. If after using single pixels the product or reference data are still not complete, it is good practice to either discard this match-up or assume constant LAI in the unmapped areas of the match-up mapping unit. If constant LAI is assumed for the reference LAI, the accuracy error of the reference data should be increased. If the unmapped area land cover is not the same as the rest of the product pixel it is good practice to use the worst case LAI range from Table 4. Alternatively, if the land cover is constant in the product pixel then the range of reference ESU LAI within the pixel should be used to model the accuracy error of the reference LAI for the pixel.

Secondly, product and reference mapping unit boundaries may not align. In this case the match-up mapping unit should ideally correspond to a spatial aggregation of sufficient product mapping units so that the relative proportion of partial reference mapping units to the match-up area is small. It is good practice to use match-up mapping units containing multiple reference mapping units. For raster datasets the match-up mapping unit could correspond to aggregations of product pixels that encompass at least 3 x 3 reference map pixels. Generalising this condition suggests the match-up mapping unit should be one order of magnitude rather than the reference mapping units.

Example 7: Geolocation Examples

Comparing a 30m resolution reference LAI raster with 1km resolution LAI products, both with 95% geolocation error of half a pixel.

The spatial accuracy of reference raster is 60m x 60m and of the product is 2km x 2km. The match-up mapping units should correspond to 3 x 3 product pixels to avoid having to bin product pixels. This area ($9km^2$) is over 10 times the area of a 30m resolution pixel, so binning error for reference pixels is not a concern. The match-ups would be the set of all reference and product pixels whose centroid falls in a moving 3km x 3km window. Every second product row and column should be used to define centroids to reduce statistical dependancies between samples. This still results in a 50% oversampling so the sample sieze used when computing statistical confidence intervals for error statistics based should correspond to half the sampled pixels.

Comparing a 60m resolution product LAI raster with no geolocation error with a 20m resolution LAI reference with a 1 pixel geolocation error.

The spatial accuracy of the product is 60m x 60m and of the reference is 100m x1 00m. Hence sampling of at least 300m x 300m ($90000m^2$) windows over the reference map grid should be used for match-ups. To avoid binning errors the match-up mapping units should be at least 10 x 60m x 60m = $36000m^2$. This means that the sampling window should be at least 189m x 189m. The 300m x 300m window is adequate.. The match ups would be the set of all reference and product pixels whose centroids fall in moving 300m x 300m windows. Every 15th reference row and column should be used to specify centroid of the sampling window giving a 50% oversampling.

Comparing an LAI 3 ESU corresponding to a 0.2ha polygon with a 20m resolution LAI products with 5m geolocation error.

This is an area of discontinuous crops with a range of LAI for the stratum is [1.0,5.0]. In this case the product suggests a match-up window size of at least 30m x 30m or here 3 x 3 pixels to avoid product binning. Reference binning error is not a concern since the reference measurement is smaller than the product measurement. Completeness is a concern since the reference measurement covers only half a pixel. Assuming LAI 3.0 over an entire product pixel could at worst case result in an +/-1.0 unit error. This should be included in the reference accuracy using a Euclidean sum. The match-up would be the LAI 3.0 and the product LAI over the 3 x 3 pixel window centred over the ESU. This is a single product mapping unit so the confidence interval of the accuracy statistic cannot be estimated.

5.2 Validation Metrics

Definitions for accuracy, precision and completeness applicable to LAI validation drawn from experimental statistics are provided here. As a good practice, validation exercises should also explicitly define these terms and identify how they relate to the definitions provided here to facilitate understanding of results across studies.

5.2.1 Definitions

These definitions are adopted from the Joint Committee for Guides in Metrology (JCGM-100 2008).

The **total measurement uncertainty** includes systematic measurement error and random measurement error. Where there is only one product estimate for each mapping unit the total measurement uncertainty corresponds to the **accuracy**.

Bias, is the expected value of the difference between corresponding product and reference estimates. Bias is an estimate of the systematic measurement error.

Precision is the dispersion of product estimates around their expected value for the same actual LAI conditions. Precision is an estimate of random measurement error.

Completeness is the proportion of valid retrievals over an observation domain.

5.2.2 Stratification of Performance Statistics

LAI validation should be performed across a representative sampling of LAI magnitudes within a spatial and temporal stratification.

LAI products are time-series maps, so complete validation should ideally include comparing spatial and temporal patterns of LAI. This involves two additional degrees of freedom over which reference samples must be acquired in addition to considering the mean LAI magnitude over a given location across time. Product precision and consistency can and should include these considerations. However, product accuracy requires reference LAI data that are and will continue to be limited until systems for high accuracy automated reference LAI mapping are developed. To avoid confusion due to differences in stratification used for accuracy, precision and completeness, it is recommended to use the constraints demanded for accuracy assessment as a single stratification.

In this regard it is good practice to:

- Employ a spatial stratification for performance assessments corresponding to continental biomes as in figure 12.
- Employ a temporal stratification e.g. separation of snow free and snow covered

conditions chiefly to recognise that most current reference LAI sampling does not consider snow covered conditions

- Sample across a representative range of LAI within a stratification for all performance statistics
- Evaluate the precision and completeness of spatial and temporal patterns in addition to reporting statistics based on LAI product estimates in a stratum without spatial or temporal considerations.

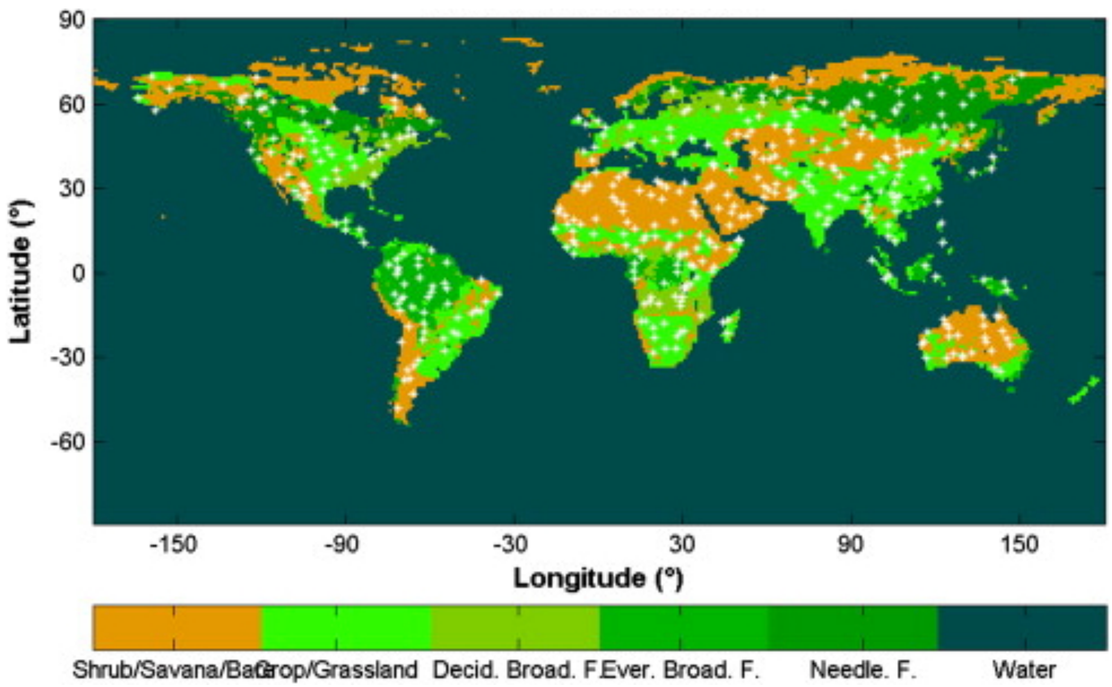


Figure 12: Strata for global LAI validation by CEOS, together with BELMANIP2 regional sites (yellow triangles). From (Weiss *et al.* Submitted).

Assessing the extent to which the sample selected conforms to these good practices is difficult since current sample sizes are not sufficiently large. A first approximation involves using the product being validated (or an ensemble of such products) to model the population of actual LAI values. In this case it is good practice to compare the cumulative frequency distribution of product LAI over the stratum and of the expected value of product LAI over match-up units. Deviations between these two distributions are indicative of limitations of the validation statistics representativeness for the stratum due to biased sampling of LAI magnitudes. For example, figure 13 compares the cumulative distribution functions of the ECOCLIMAP LAI product for the BELMANIP v1 (Baret *et al.* 2006) sampling design with an exhaustive global sampling. The comparison suggests that BELMANIP v1 under-samples bare regions but provides representative sampling between LAI 1.0 and 4.0.

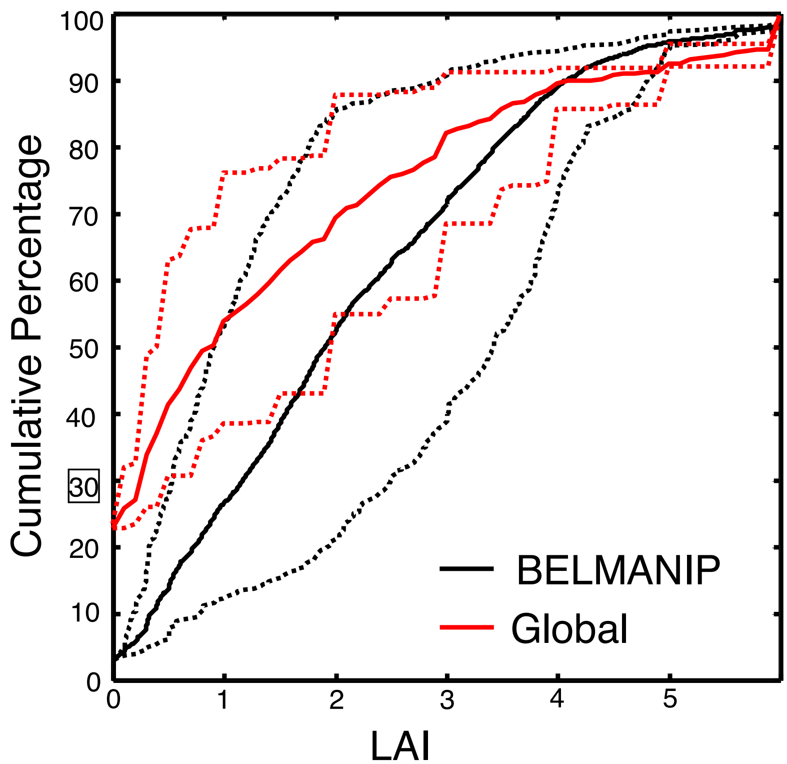


Figure 13: Comparison of the cumulative distribution of the ECOCLIMMAP peak season LAI within the BELMANIP sampling design over a global extent (all mapped land pixels). From (Baret *et al.* 2006).

5.2.3 Validation Statistics

Validation statistics should be reported for each stratum. However, the sampling distribution will vary between accuracy, precision and completeness. It is good practice to define the sampling distribution and samples from which validation statistics are derived. This section specifies good practice statistics or visualisation of performance. In addition we relate this good practice to the commonly used practice in the current literature. For reference, table 6 summarises both the common practice and the recommended good practice.

Table 6: Recommended validation statistics.

Category	Good Practice	Current Practice
Total Measurement Error	Scatter plot of mean or median match-ups	Scatter plot of mean only match-up
	Median and percentiles of absolute residuals, RMSE	Root Mean Square Error (RMSE)
	Box plot of absolute residuals vs LAI	Scatter plot of residuals versus LAI
Bias	Median and percentiles of residuals	Mean difference
	Box plot of residuals vs. LAI	Mean difference vs. LAI
	Kendall-Thiel line slope and confidence interval	Ordinary least squares line slope and confidence interval
Precision	Box plot of residuals from Kendall-Thiel Line fit	Residuals of line fit versus LAI.
	Median signed anomaly of 95th percentile and 5th percentile	Mean seasonal difference
	Median 3 point difference	Mean 3 point difference
	Spatial rank correlation	Pearson's correlation coefficient
Completeness	Gap size distribution	Relative frequency histograms

5.2.3.1 Measurement Uncertainty

Statistics related to measurement uncertainty is derived from the reference and product samples within each match-up in a stratum. Match-ups will typically have more than one reference or product value. For simplicity, the expected value of differences (residuals) within a match-up is used for uncertainty statistics. When a single ESU reference value, is used the expected value is the median of all pairwise residuals. Otherwise the expected value is the difference in the mean of reference and product values. A scatter plot of expected values of product versus reference LAI for all match-ups should be reported together with a 1:1 line (e.g. figure 14).

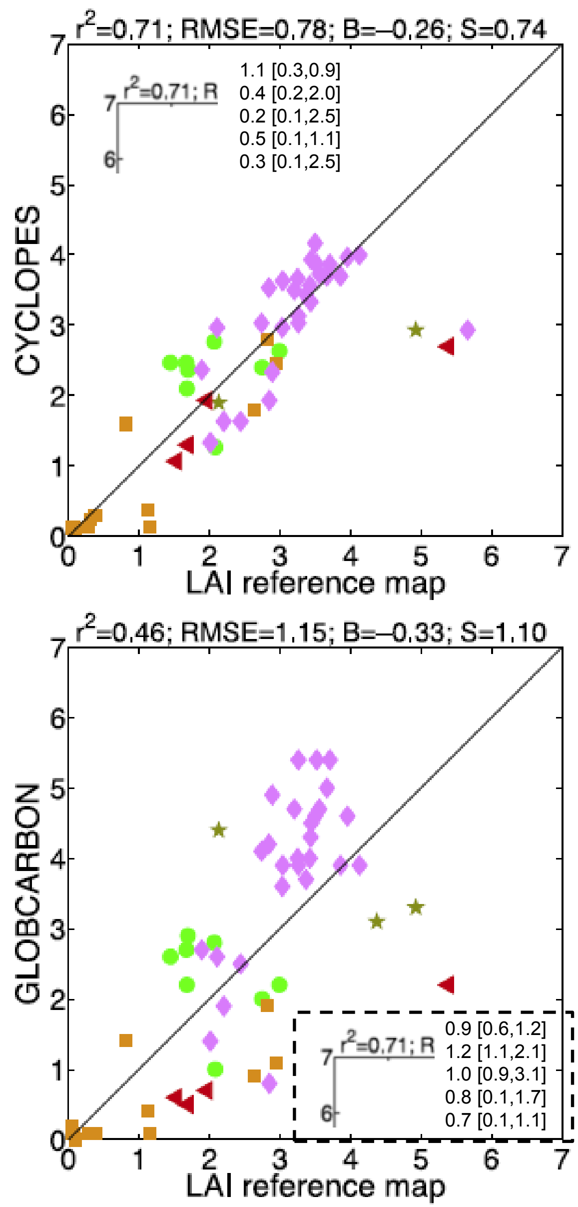


Figure 14: Scatter plots between reference and product LAI for a global validation study together with the median absolute difference and range of absolute differences as a function of biome. From (Garrigues *et al.* 2008a).

Under the assumption that the distribution of residuals between match-ups is unimodal, it is good practice to report the median absolute difference as the total measurement uncertainty for this stratum. To represent the possibility of outliers or multiple modes accuracy could be further characterised by plotting absolute differences against their percentile (e.g. figure 11). In cases where there are sufficient direct comparisons a density plot of the product and reference LAI should be reported (e.g. figure 15).

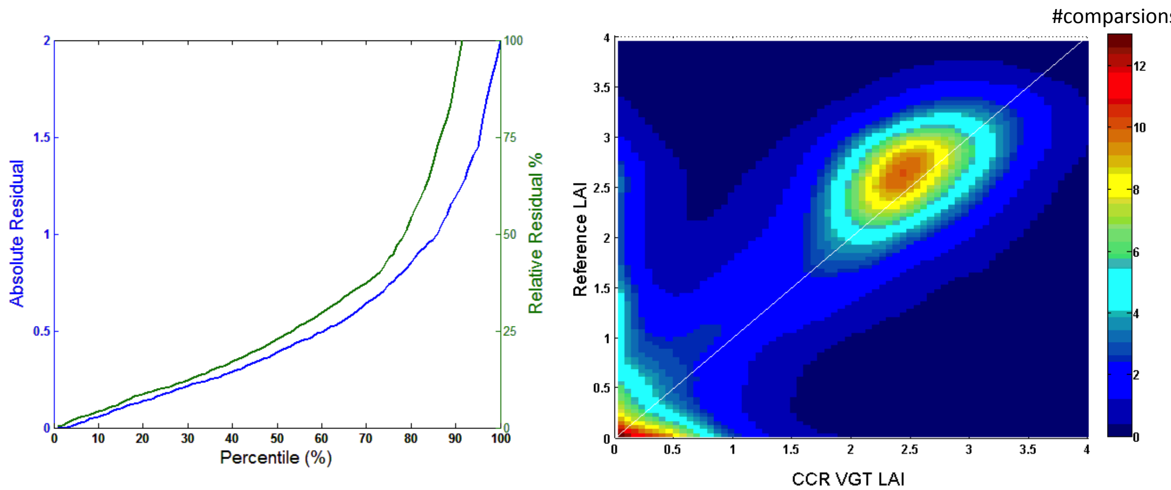


Figure 15: Comparison of CCRS SPOT VGT LAI product with the reference LAI map shown in Example 5. Left panel shows the percentile rank of both absolute residual (blue line) and relative residuals (green line). Right panel provides a bivariate density plot.

LAI products often show biases in residuals as a function of product LAI. Firstly, for products derived using passive optical measurements an asymptote frequently exists at high (>3.0) LAI irrespective of the underlying true LAI. Secondly, a linear bias as a function of LAI may be present due to a constant error in canopy clumping. It is good practice to product a scatter plot of the signed residuals as a function of product LAI and to summarise this relationship numerically. For example, figure 16 summarises the absolute residuals (the median indicates total measurement uncertainty), the signed residual (the median indicates bias) and the residual after linear regression of reference versus product LAI (the median indicating precision) as a function of LAI. In this example, residuals are <1.0 until LAI 6.0. Removal of linear trends in residuals does not substantially reduce their magnitude, indicating the majority of residuals are random in nature (below LAI 6.0) or due to saturation of the product retrievals (LAI >6.0). A spatially explicit map of accuracy error should be reported for products that span multiple reference datasets.

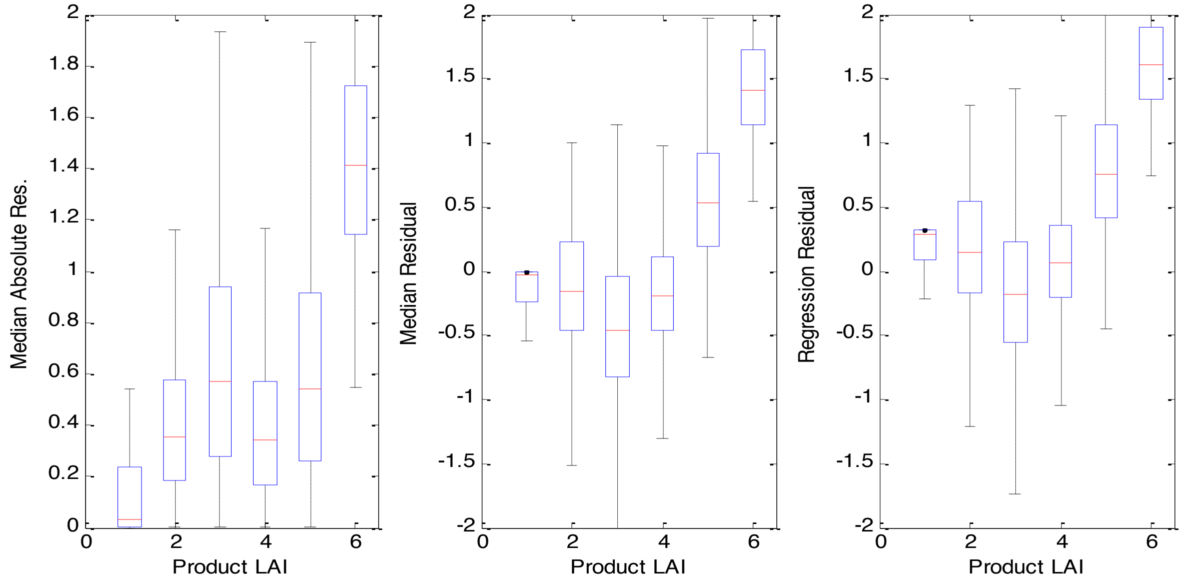


Figure 16: Box-plots of measurement uncertainty statistics from comparison of CCRS VGT LAI and reference LAI map shown in Example 5. Red bars indicate median residuals, blue boxes cover 50% of the data and whiskers include 95% of the data.

5.2.3.2 Precision

Precision corresponds to the dispersion of an ensemble of product LAI estimates. Ideally the precision of both the spatial and temporal evolution of LAI should be quantified. This is challenging to visualise and can be sensitive to disturbances (e.g. forest fires and droughts) or shifts in the phase of seasonal LAI cycles. We assume that spatial disturbances do not impact the majority of the stratum (this should be evaluated if possible). We also assume that temporal shifts do not impact the seasonal range of LAI for the majority of annual cycles (this should be evaluated if possible). With these assumptions three aspects of precision are proposed for evaluation.

a) Inter-annual Precision

Anomalies of an upper and lower percentile of LAI are indicators of intern-annual precision, for example the lower 5th percentile and upper 95th percentile of annual LAI. In the absence of a linear trend the dispersion of these anomalies about their expected value is then a useful indicator of inter-annual temporal precision. Ideally a Mann-Kendall test could be applied to detect and remove linear trends although this is not recommended considering the relatively short (<30 years) temporal extent of most products.

For all pixels without (or assumed to be without) detected linear trends in anomalies it is good practice to report the median absolute deviation of anomalies and the confidence interval of this statistic as spatial maps. As a non-spatial statistic, it is good practice to provide a boxplot of the median absolute deviation of anomalies versus product LAI for bins corresponding to LAI 0.0 – 1.0, 1.0 – 2.0, 2.0 – 3.0 etc. until the maximum product LAI together with a single statistic corresponding to the 50th percentile of anomalies over the entire stratum.

b) Intra-annual Precision

Intra-annual precision corresponds to temporal noise assumed to have no serial correlation within a season. In most strata we can assume that the actual LAI between retrievals separated by a short (e.g. 10 day or less) time period should change at a rate similar to a slightly longer (e.g. 30 day) period. In this case the anomaly of a product LAI value from the linear estimate based on its neighbours can be used as an indication of intra-annual precision. Figure 17 provides an example of global histograms of central differences. It is possible that intra-annual noise will vary with biome and position in the seasonal cycle. To account for this possibility it is good practice that the observed anomalies for each month and each biome are summarised in terms of box plots as a function of the seasonal cycle of LAI.

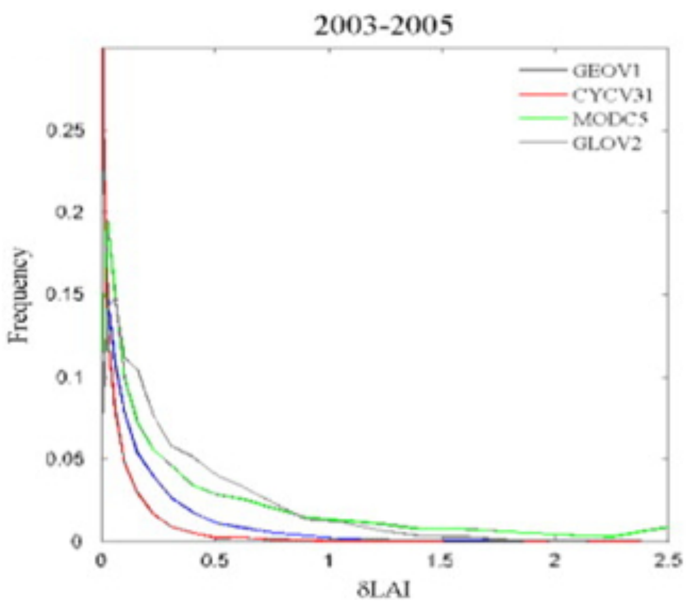


Figure 17: Historgam of differences for four global LAI products over all BELMANIP2 sites for 3 years (Camacho *et al.* 2011).

GEOV1 (GEOLAND v1), CYCV31 (CYCLOPES v3.1), MODC5 (MODIS Collection 5) and GLOV2 (GLOBCARBON v2).

c) Spatial Precision

Spatial precision corresponds to the repeatability of the spatial pattern of a product. Spatial rank correlation can be used to quantify the pairwise similarity of two products from the point in the annual cycle while minimising the impact of temporal noise. It is good practice that the rank correlation between successive temporal products should be computed for every annual cycle and summarised in terms of box plots as a function of the seasonal cycle of LAI. Rank correlation statistics are useful to evaluate product performance over time but should not be used to compare the precision of products with substantially different spatial mapping unit size.

5.2.3.3 Completeness

Completeness corresponds to the absence of spatial or temporal gaps in data. It is good practice to map the proportion of good retrievals within a stratum. Spatial completeness will change with season. Figure 18 shows both temporal and spatial completeness using multiple years of global LAI products. Temporal completeness is depicted using the monthly average percentage of valid retrievals. Spatial completeness uses bar plots to summarise percentage valid retrievals by biome. To reflect the potential variability in temporal completeness from year to year and biome to biome it is good practice to provide a box plot for completeness based on averages taken for each unique year and biome condition and to provide biome specific plots for outlier biomes. Similarly, it is good practice to report spatial completeness as the box plots of the proportion of good retrievals as a function of the seasonal cycle of LAI. Temporal completeness includes consideration of the persistence of gaps. It is good practice to graph the frequency of a given temporal gap size versus the temporal gap size over the available data (e.g. figure 19).

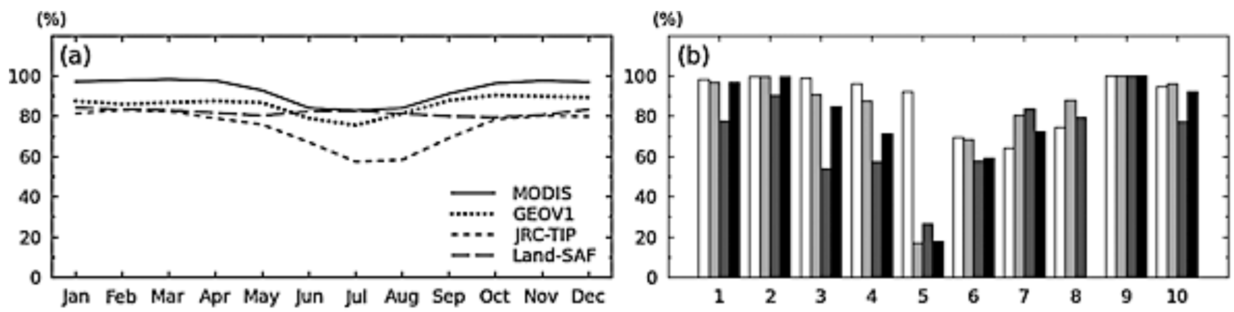


Figure 18: Percentage of acceptable quality retrievals (according to producer provided quality indices) as a function of time (a) or biome (b). Note that these statistics assume accurate data quality reporting by producers. From (Fang *et al.* 2013).

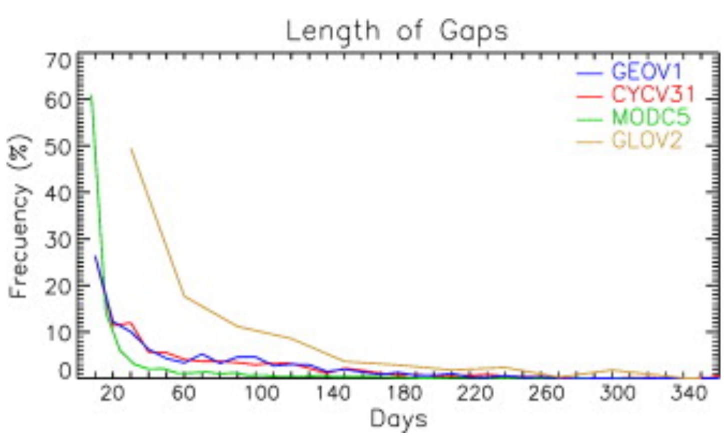


Figure 19: Gap length frequency for four Global LAI products over all BELMANIP2 sites for 3 years (Camacho *et al.* 2012).

GEOV1 (GEOLAND v1), CYCV31 (CYCLOPES v3.1), MODC5 (MODIS Collection 5) and GLOV2 (GLOBCARBON v2).

5.2.3.4 Ensemble Inter-comparison

Ensemble LAI estimates from multiple products offer a means of simultaneously evaluating precision and stability. In the ideal case a sufficient number of independent ensemble members (products) would be available to quantify the confidence intervals for these statistics. At present this is not the case so the good practices for quantifying confidence intervals of inter-comparisons have yet to be developed and tested.

Current examples of good practice for ensemble inter-comparison involve one of two procedures:

1. Mapping residuals between ensemble members and some candidate reference from the ensemble (e.g. figure 20).
2. Comparing the joint distribution (scatter plots, e.g. figure 21) or marginal distributions (histograms, e.g. figure 22) of LAI of ensemble members with the reference within a stratum.

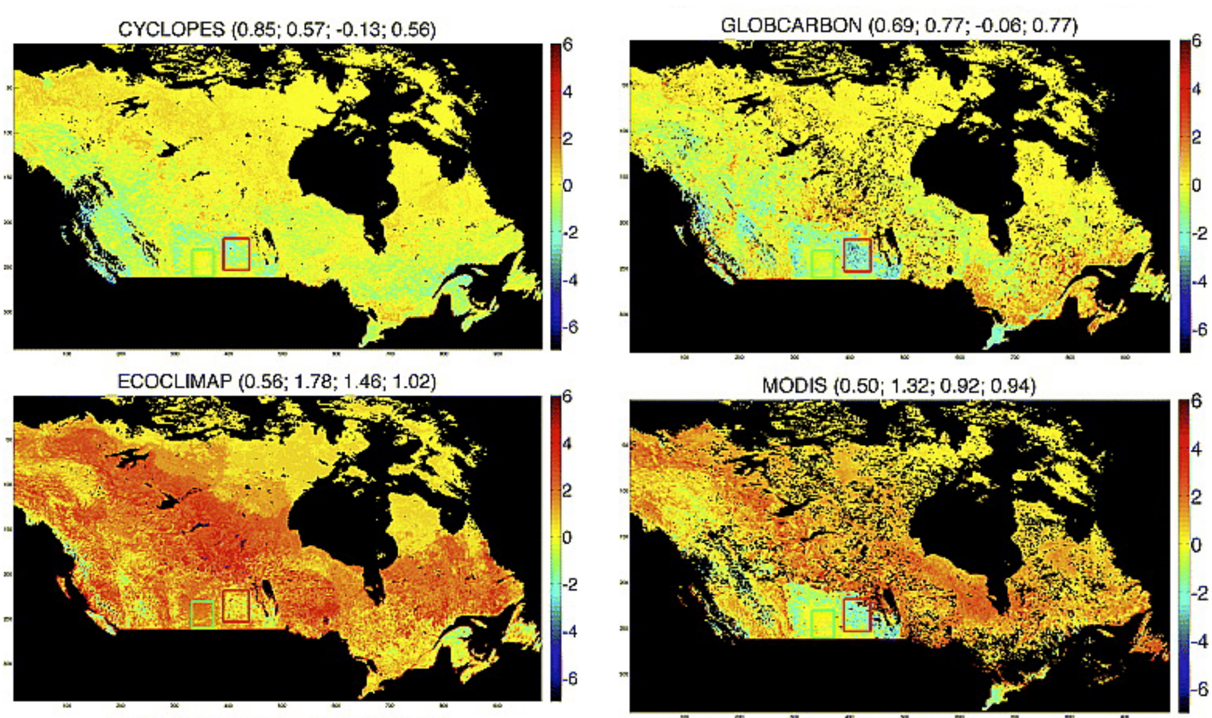


Figure 20: Anomalies in peak season LAI between global products and a chosen reference member (in this case a regional LAI product over Canada and Alaska). Boxes indicate regions with differences related to land cover specification in global products. From (Garrigues *et al.* 2008a).

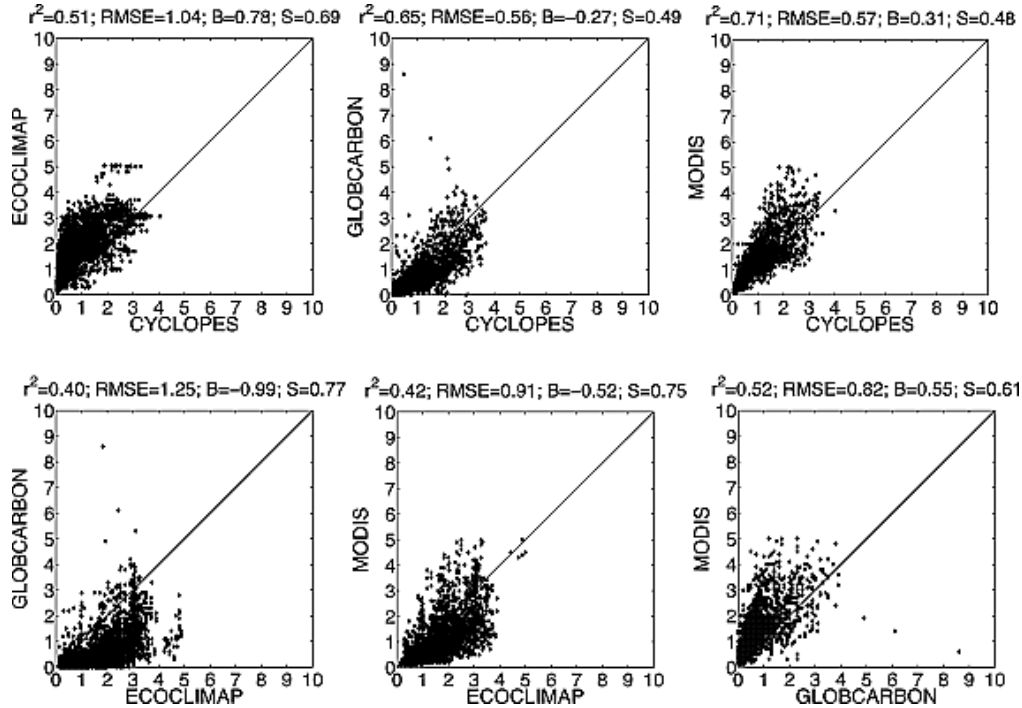


Figure 21: Scatter plots comparing growing season LAI for four global LAI products over grassland biome BELMANIP sites for four years. Included are summary statistics for linear correlation coefficient (R^2), root mean square error (RMSE), mean signed bias (B), and standard deviation of differences (S). From (Garrigues *et al.* 2008a).

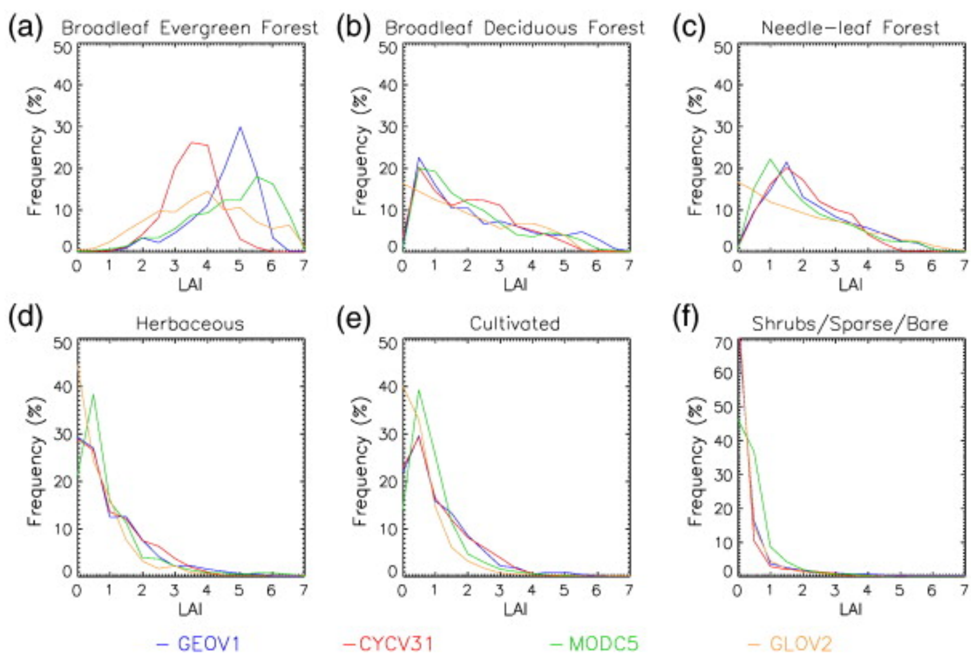


Figure 22: Inter-comparison of product LAI histograms on a biome basis. From (Camacho *et al.* 2011).

Both comparisons should be performed at a common aggregation level both in space and time. Current good practice relies chiefly on monthly temporal intervals and spatial resolutions on the order of 10km x 10km (e.g. (Garrigues *et al.* 2008a, Fang *et al.* 2013)). With the advent of higher spatial resolution systematic LAI products it is anticipated that these good practices will require adjustment. One good practice would be to use the same rules for defining match-up mapping units as identified for validation in Section 5.1.3.

It is proposed that irrespective of the aggregation used for a specific validation study an inter-comparison of products also be conducted at 10km x 10 km resolution for 3 monthly intervals spanning each year. This will ensure some level of consistency across studies. To further promote consistency, scatter plots stratified by continental biomes should be reported where feasible.

The possibility for using triple-co-location (Stoffelen 1998) for ensemble inter-comparison needs further investigation before it can be recommended as a good practice. Chiefly, a number of LAI products are derived from the same input satellite imagery or trained from other LAI products. In both instances the assumption that each product is an independent estimate of the underlying LAI may not apply so co-location methods may provide biased precision and agreement statistics.

5.3. Reporting Results of LAI Validation

The results of validation exercises should be reported publicly after review by the data producers and independent scientific peer review. Reporting in refereed journals is encouraged and supporting materials corresponding to spatial or temporal accuracy statistics should be made accessible. The following details related to reporting are good practices:

1. All participants in the exercise should be declared unless products were provided blindly.
2. Links to accessible versions of the products and reference data used during the validation should be provided and maintained.
3. Match-ups of product and reference LAI values used to derive aggregation statistics together with ancillary information related to location (at least the continent and biome), temporal interval (at least snow or snow free condition) and uncertainty in reference data (at least a reference to the protocol used to produce each reference data point) should be made available publicly.
4. Scatter plots and statistics recommended in Section 5.2 and 5.3 should be reported within the validation document or linked supplementary material in addition to any other statistics.
5. Planned updates or revisions to the document (e.g. in anticipation of new reference datasets that may be available on a regular basis) should be identified.

6 CONCLUSIONS

A number of good practices for validation of global leaf area index products are presented within this document. The good practices include essential definitions, an assessment of approaches for producing reference datasets, the current status of validation and a recommended validation methodology based on current knowledge.

The validation methodology involves quantification of product accuracy, precision and completeness. Globally representative sampling schemes, such as BELMANIP2, should be used to prioritise new in situ LAI networks. However, the good practices presented recognise that this may not be feasible with current LAI survey methods. Two strategies were proposed to address this limitation. Firstly, the use of robust biome level statistics for reporting product accuracy so that confidence in our knowledge of global product accuracy is fairly represented. Secondly, the use of upscaling methods that maximise the coverage of reference LAI maps. Specifically, careful consideration of in situ sampling and design of transfer functions for upscaling can produce reference maps with relatively low bias over large (>100km^2) regions.

The good practices promote uniform definitions for vegetation quantities related to LAI, uniform spatial and temporal aggregation procedures, sampling design for reference comparisons and inter-comparison and uniform performance statistics and visualisations. These approaches should be considered a minimum common element of validation studies and should be enhanced where feasible with performance indicators relevant to each datasets available during a validation study. New validation exercises are encouraged to implement these metrics and report to CEOS LPV in terms of their information content.

While this document serves to summarise a common set of knowledge and methods useful for validating the mapped satellite-derived LAI values in a product, it does not include best practices related to explaining observed LAI errors. There is a need to develop a traceable quality assurance system for evaluating in situ methodologies and satellite retrieval algorithms under controlled conditions to relate observed differences between product and reference LAI to deficiencies in algorithms, satellite data or potential reference datasets.

7 REFERENCES

Asner, G. P., Scurlock, J. M. O. and Hicke, J. A. (2003). Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecology and Biogeography.* **12**(3): 191-205.

Barclay, H. J. (1998). Conversion of total leaf area to projected leaf area in lodgepole pine and Douglas-fir. *Tree Physiology.* **18**(3): 185-193.

Baret, F. (2012). VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite products. http://w3.avignon.fr/valeri/documents.

Baret, F., de Sloan, B., Lopez-Lozano, R., Ma, K. and Weiss, M. (2010). LAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5degree zenith angle. *Agricultural and Forest Meteorology.* **150**: 1393-1401.

Baret, F. and Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment.* **35**: 161–173.

Baret, F., Morisette, J., Fernandes, R., Champeaux, J., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S. and Nickeson, J. (2006). Evaluation of the representativeness of network of sites for the Global Validation and Intercomparison of Land Biophysical Products. *IEEE Transactions on Geoscience and Remote Sensing.* **44**: 1794-1803.

Breda, N. J. (2003). Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany.* **54**(392): 2403-2417.

Brown, L., Chen, J. M., Leblanc, S. G. and Cihlar, J. (2000). A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote Sensing of Environment.* **71**(1): 16-25.

Camacho, F., Cemicharo, J., Lacaze, R., Baret, F. and Weiss, M. (2012). GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. *Remote Sensing of Environment.* **137**: 310-329.

Camacho, F. and Cernicharo, J. (2010). Validation Report Low Resolution (SPOT/VGT) Vegetation Variables (GEOV1). BP-RP-BP022 Geoland 2. http://web.vgt.vito.be/documents/BioPar/g2-BP-RP-BP022-ValidationReport-VGTV1Products-I1.00.pdf.

Camacho, F. and Cernicharo, J. (2011). Validation Report Medium Resolution (MERIS) Vegetation Parameters. BP-RP-BP022. Geoland 2. pp62. http://web.vgt.vito.be/documents/BioPar/g2-BP-RP-BP022-ValidationReport-MERISProducts-I1.00.pdf.

Chen, J. M. and Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant Cell and Environment.* **15**: 421-429.

Chen, J. M. and Cihlar, J. (1995). Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index. *Applied Optics.* **34**(27): 6211-6222.

Chen, J. M., Govind, A., Sonnentag, O., Zhang, Y. Q., Barr, A. and Amiro, B. (2006). Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology.* **140**(1-4): 257-268.

Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S. G., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymow, J. A., Swift, E., Van der Sanden, J. and Pellikka, P. K. E. (2002). Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sensing of Environment.* **80**(1): 165-184.

Chen, J. M., Rich, P. M., Gower, S. T., Norman, J. M. and Plummer, S. (1997). Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical*

Research-Atmospheres. **102**(D24): 29429-29443.
Chen, W., Li, J., Zhang, Y., Zhou, F., Koheler, K., Leblanc, S., Fraser, R., Olthof, I. and Wang, J. (2009). Relating biomass and leaf area index to non-destructive measurements for monitoring changes in arctic vegetation. *Arctic.* **62**: 281-294.
Cheng, C. L. and Van Ness, J. (1999). *Statistical Regression with Measurement Error*. London, Arnold Publishers. pp269.
Claverie, M., Vermote, E., Weiss, M., Baret, F., Hagolle, O. and Demarez, V. (2013). Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France. *Remote Sensing of Environment.* **139**: 216-230.
Cohen, W. B., Maiersperger, T. K., Turner, D. P., Ritts, W. D., Pflugmacher, D., Kennedy, R. E., Kirschbaum, A., Running, S. W., Costa, M. and Gower, S. T. (2006). MODIS land cover and LAI collection 4 product quality across nine sites in the western hemisphere. *IEEE Transactions on Geoscience and Remote Sensing.* **44**(7): 1843-1857.
Cutini, A. (2002). Litterfall and Leaf Area Index in the CONECOFOR Premanent Monitoring Plots. *Journal of Limnology.* **61**: 62-68.
Decagon (2012). Sunflex PAR Cemptometer, Operators Manual, Decagon technical report. http://www.decagon.com/assets/Manuals.
Demarez, V., Duthoit, S., Baret, F., Weiss, M. and Dedieu, G. (2008). Estimation of leaf area and clumping indexes of crops with hemispherical photographs. *Agricultural and Forest Meteorology.* **148**(4): 644-655.
Duda, R., Hart, P. and Stork, D. G. (2000). *Pattern Classification*. 2nd ed. Wiley. pp637.
ESA (2007). GLOBCARBON DPQR v4.2 demonstration products and qualification report. ESA. pp69. http://due.esrin.esa.int/prjs/prjs43.php.
Fang, H. L., Jiang, C. Y., Li, W. J., Wei, S. S., Baret, F., Chen, J. M., Garcia-Haro, J., Liang, S. L., Liu, R. G., Myneni, R. B., Pinty, B., Xiao, Z. Q. and Zhu, Z. C. (2013). Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties. *Journal of Geophysical Research-Biogeosciences.* **118**(2): 529-548.
Fernandes, R., Butson, C., Leblanc, S. and Latifovic, R. (2003). Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data. *Canadian Journal of Remote Sensing.* **29**(2): 241-258.
Fernandes, R. and Leblanc, S. (2005). Appropriate linear regression techniques for the calibration of remote sensing models: When classical linear regression should not be used *Remote Sensing of Environment.* **95**(3): 303-316.
Fernandes, R. A. (2012). Canada Centre for Remote Sensing Protocol for In situ Leaf Area Index using Digital Hemispherical Photography using the INRA CANEYE Analysis System. CCRS Technical Report Series. ftp://ftp.ccrs.nrcan.gc.ca/EMS/LAI/CCRSCaneye.pdf.
Futmon (2009). Field protocol on Radiation measurements and Leaf Area Index. FUTMON Technical Report. www.futmon.org/sites/default/files.
Ganguly, S., Nemani, R. R., Zhang, G., Hashimoto, H., Milesi, C., Michaelis, A., Wang, W. L., Votava, P., Samanta, A., Melton, F., Dungan, J. L., Vermote, E., Gao, F., Knyazikhin, Y. and Myneni, R. B. (2012). Generating global Leaf Area Index from Landsat: Algorithm formulation and demonstration. *Remote Sensing of Environment.* **122**: 185-202.
Garrigues, S., Allard, D., Baret, F. and Weiss, M. (2006). Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sensing of Environment.* **103**: 81-96.
Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y. and Yang, W. (2008a). Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. *Journal of Geophysical Research-Biogeosciences.* **113**(G2).
Garrigues, S., Shabanov, N., Sweanson, K., Morisette, J. T., Baret, F. and Myneni, R. B. (2008b). Intercomparison and sensitivity analysis of Leaf Area Index retrievals from LAI-

2000, AccuPAR and digital hemispherical photography over croplands. 16. NASA. http://digitalcommons.unl.edu/nasapub/16.
GCOS-92 (2004). Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC. WMO/TD No. 1219. GCOS. http://www.wmo.int/pages/prog/gcos/Publications/gcos-92_GIP.pdf.
GCOS-107 (2006). Systematic Observation Requirements for Satellite-based Products for Climate: Supplemental Details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in support of the UNFCCC (GCOS-92)". WMO/TD No. 1338. GCOS. http://www.wmo.int/pages/prog/gcos/Publications/gcos-107.pdf.
GCOS-138 (2010). Implementation Plan for the Global Observing System for Climate in support to UNFCCC (2010 Update). WMO-TD/No. 1523. GCOS. http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf.
Gobron, N. and Verstraete, M. M. (2009). Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables: Leaf Area Index (LAI). Global Terrestrial Observing System. pp12. http://www.fao.org/gtos/doc/ECVs/T11/T11.pdf.
Good, P. I. and Hardin, J. W. (2009). *Common Errors in Statistics (And How to Avoid Them)*. 3rd ed. Hoboken, New Jersey, Wiley. pp211.
Gower, S. T., Vogel, J. G., Norman, J. M., Kucharik, C. J., Steele, S. J. and Stow, T. K. (1997). Carbon distribution and aboveground net primary productivity in aspen, jack pine, and black spruce stands in Saskatchewan and Manitoba. *Geophysical Research Atmospheres.* **102**(D24): 29029–29041.
Hastie, T., Tibshirani, R. and Friedman, J. (2011). *The Elements of Statistical Learning*. Springer. pp729.
Heiskanen, J., Rautiainen, M., Korhonen, L., Mottus, M. and Stenberg, P. (2011). Retrieval of boreal forest LAI using a forest reflectance model and empirical regressions. *International Journal of Applied Earth Observation and Geoinformation.* **13**(4): 595-606.
Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory.* **14**(1): 55-63.
JCGM-100 (2008). Evaluation of measurement data — Guide to the expression of uncertainty in measurement http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.
Jonckherre, I., Fleck, S., Nackerts, K., Muysa, B., Coppin, P., Weiss, M. and Baret, F. (2004). Review of methods for in sit leaf area index determination Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology.* **121**: 19-35.
Kucharik, C. J., Norman, J. M. and Gower, S. T. (1998). Measurements of branch area and adjusting leaf area index indirect measurements. *Agricultural and Forest Meteorology.* **91**(1-2): 69-88.
Lang, A. and Xiang, Y. (1986). Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies. *Agricultural and Forest Meteorology.* **19**: 187-207.
Law, B. E., Arkebauer, T., Campbell, J. L., Chen, J., Sun, O., Schwartz, M., van Ingen, C. and Verma, S. (2008). Terrestrial Carbon Observations: Protocols for Vegetation Sampling and Data Submission. Report 55. Global Terrestrial Observing System, FAO. pp87.
Leblanc, S. (2005c). Tracing Radiation and Architecture of Canopies Manual. CCRS Technical Report. ftp://ftp.ccrs.nrcan.gc.ca/ad/LEBLANC/SOFTWARE/DHP/TRAC_MANUAL.pdf.
Leblanc, S. (2008). DHP-TRACWin Manual. CCRS Technical Report. ftp://ftp.ccrs.nrcan.gc.ca/ad/LEBLANC/SOFTWARE/DHP/DHP-TRACWIN_MANUAL.pdf.
Leblanc, S. G., Chen, J. M., Fernandes, R., Deering, D. W. and Conley, A. (2005a). Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology.* **129**(3-4): 187-207.
Licor (2010). LAI-2200 Instruction Manual. 984-10633. LI-Cor Inc.

Liu, J. G., Pattey, E. and Admiral, S. (2013). Assessment of in situ crop LAI measurement using unidirectional view digital photography. *Agricultural and Forest Meteorology.* **169**: 25-34.

Macfarlane, C. (2011). Classification method of mixed pixels does not affect canopy metrics from digital images of forest overstorey. *Agricultural and Forest Meteorology.* **151**(7): 833-840.

Majasalmi, T., Rautiainen, M., Stenberg, P. and Rita, H. (2012). Optimizing the sampling scheme for LAI-2000 measurements in a boreal forest. *Agricultural and Forest Meteorology.* **154**: 38-43.

Malone, S., Herbert, D. A. and Holshouser, D. L. (2002). Evaluation of the LAI-2000 plant canopy analyzer to estimate leaf area in manually defoliated soybean. *Agronomy Journal.* **94**(5): 1012-1019.

Martinez, B., Cassiraga, E., Camacho, F. and Garcia-Haro, J. (2010). Geostatistics for Mapping Leaf Area Index over a Cropland Landscape: Efficiency Sampling Assessment. *Remote Sensing.* **2**(11): 2584-2606.

Miller, J. B. (1967). A formula for average foliage density. *Australian Journal of Botany.* **25**: 142-144.

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R. A., Leblanc, S. G., Kalacska, M., Sanchez-Azofeifa, G. A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A. N., Lewis, T. E., Iiames, J. S., Colombo, R., Meroni, M., Busetto, L., Cohen, W. B., Turner, D. P., Warner, E. D., Petersen, G. W., Seufert, G. and Cook, R. (2006). Validation of global moderate-resolution LAI products: A framework proposed within the CEOS Land Product Validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing.* **44**(7): 1804-1817.

NEON (2009). The Neon Strategy. pp40. http://www.neoninc.org/sites/default/files/NEON.Strategy.July2009.Release2_2_0.pdf.

Nilson, T. (1971). A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meterology.* **8**: 25-38.

Nilson, T. (1999). Inversion of gap frequency data in forest stands. *Agricultural and Forest Meteorology.* **98-9**: 437-448.

Nilson, T. and Kuusk, A. (2004). Improved algorithm for estimating canopy inidces from gap fraction data in forest canopies. *Agricultural and Forest Meteorology.* **124**: 157-169.

Oker-Blom, P. and Smolander, H. (1988). The ratio of shoot silhouette area to total needle area in Scots pine. *Forest Science.* **34**: 894-906.

Pekin, B. and Macfarlane, C. (2009). Measurement of Crown Cover and Leaf Area Index Using Digital Cover Photography and Its Application to Remote Sensing. *Remote Sensing.* **1**(4): 1298-1320.

Pinty, B., Jung, T., Kaminski, T., Lavergne, M., Mund, M., Plummer, S., Thomas, E. and Widlowski, J.-L. (2011). Evaluation of the JRC-TIP 0.01 degree products over a mid-latitude deciduous forest site. *Remote Sensing of Environment.* **115**(12): 3567 - 3581.

Pitman, R., Bastrup-Birk, A., Breda, N. and Rautio, P. (2010). Sampling and analysis of Litterfall. In: Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests. UNECE ICP Forests Progamme Co-ordinating Center. pp16. http://www.icp-forests.org/Manual.htm.

Plummer, S., Arino, O., Simon, M. and Steffen, W. (2006). Establishing a earth observation product service for the terrestrial carbon community: The GLOBCARBON initiative. . *Mitigation and Adaptation Strategies for Global Change.* **11**: 97–111.

Privette, J. L., Myneni, R. B., Knyazikhin, Y., Mukelabai, M., Roberts, G., Tian, Y., Wang, Y. and Leblanc, S. G. (2002). Early spatial and temporal validation of MODIS LAI product in the Southern Africa Kalahari. *Remote Sensing of Environment.* **83**(1-2): 232-243.

Ryu, Y., Sonnentag, O., Nilson, T., Vargas, R., Kobyashi, H., Wenk, R. and Baldocchi, D. (2010). How to quantify tree leaf area index in an open savannah ecosystem. *Agricultural and Forest Meteorology.* **150**: 63-76.

Ryu, Y., Verfaillie, J., Macfarlane, C., Kobayashi, H., Sonnentag, O., Vargas, R., Ma, S. and Baldocchi, D. D. (2012). Continuous observation of tree leaf area index at ecosystem scale using upward-pointing digital cameras. *Remote Sensing of Environment.* **126**: 116-125.

Schowengerdt, R. A. (1997). *Remote Sensing Models and Methods for Image Processing*. Academic Press. pp560.

Shabanov, N. V., Huang, D., Yang, W. Z., Tan, B., Knyazikhin, Y., Myneni, R. B., Ahl, D. E., Gower, S. T., Huete, A. R., Aragao, L. E. O. C. and Shimabukuro, Y. E. (2005). Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests. *Ieee Transactions on Geoscience and Remote Sensing.* **43**(8): 1855-1865.

Stenberg, P. (1996). Correcting LAI-2000 estimates for the clumping of needles in shoots of conifers. *Agricultural and Forest Meteorology* **79**: 1-8.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research-Oceans.* **103**(C4): 7755-7766.

Verger, A., Baret, F. and Camacho, F. (2011). Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. *Remote Sensing of Environment.* **115**(2): 415-426.

Weiss, M. (2002). VALERI: Deriving leaf area index and average leaf inclination angle from LAI2000 Measurements. INRA Technical Report. http://w3.avignon.inra.fr/valeri/fic_htm/methodology/main.php.

Weiss, M., Baret, F., Block, T., Koetz, B., Burini, A., Scholze, B., Brockmann, C., Fernandes, R., Plummer, S., Myneni, R., Gobron, N., Nightingale, J., Schaepman-Strub, G., Camacho, F. and Sanchez-Azofeifa, A. (Submitted). On Line Validation Exercise (OLIVE): a web based service for the validation of medium resolution land products. Application to FAPAR products. *Biogeosciences.*

Weiss, M., Baret, F., Smith, G. J., Jonckheere, I. and Coppin, P. (2004). Review of methods for in situ leaf area index (LAI) determination Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology.* **121**(1-2): 37-53.

Welles, J. M. and Norman, J. M. (1991). Instrument for indirect measurement of canopy architecture. *Agronomy Journal.* **83**: 818-825.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications.* **9**(6): 476-487.

Widlowski, J.-L., Taberner, M., Pinty, B., Bruniquel-Pinel, V., Disney, M., Fernandes, R., Gastellu-Etchegorry, J.-P., Gobron, N., Kuusk, A., Lavergne, T., Leblanc, S., Lewis, P., Martin, E., Mottus, M., North, P., Qin, W., Robustelli, M., Rochdi, N., Ruiloba, R., Soler, C., Thompson, R., Verhoef, W., Verstraete, M. and Xie, D. (2007). The third RAdiation transfer Model Intercomparion (RAMI) exercise: Documenting progress in canopy reflectance modelling. *Journal of Geophysical Research.* **112**: DOI: 10.1029/2006JD007821.

Yang, B., Tan, B., Huang, D., Rautainen, M., Shabanov, N. and Wang, Y. (2006). MODIS Leaf Area Index products: from validation to algorithm improvement. *Transactions of Geoscience and Remote Sensing.* **44**: 1885-1898.

Zhang, Y., Chen, J. M. and Miller, J. R. (2005). Determining digital hemispherical photograph exposure for leaf area index determination. *Agricultural and Forest Meteorology.* **113**: 166-181.

8 APPENDIX A

Appendix A: Table 3 Methods URL Links. This table provides full URLs for the LAI methods in Table 3 of the main text for printed versions of this table. See main text for sampling type, landcover type, and citation information for each.

Name	Link
AAFC	http://www.sciencedirect.com/science/article/pii/S0168192310002029
BIGFOOT	http://www.fsl.orst.edu/larse/bigfoot/overview.html
BOREAS Destructive	http://onlinelibrary.wiley.com/doi/10.1029/97JD02317/abstract
BOREAS Non-Destructive	http://onlinelibrary.wiley.com/doi/10.1029/97JD01107/abstract
CCRS CANEYE	http://lpvs.gsfc.nasa.gov/PDF/CCRCANEYELAIv2012.pdf
CCRS DHP	ftp://ftp.ccrs.nrcan.gc.ca/ad/LEBLANC/SOFTWARE/DHP/DHP-TRACWin_MANUAL.pdf
CCRS TRAC	ftp://ftp.ccrs.nrcan.gc.ca/ad/LEBLANC/SOFTWARE/DHP/TRAC_MANUAL.pdf
CCRS Tundra	http://pubs.aina.ucalgary.ca/arctic/Arctic62-3-281.pdf
CONECOFOR	http://www.jlimnol.it/index.php/jlimnol/article/view/349/0
DECAGON Ceptometer	http://manuals.decagon.com/Manuals/10242_Accupar LP80_Web.pdf
FLUXNET	http://faculty.geog.utoronto.ca/Chen/Chen's homepage/PDFfiles/unp115_Jing6_AFM.pdf
FUTMON	http://www.futmon.org/sites/default/files/documenten/Field_Protocol_Radiation_LAI_D2_3f.pdf
GTOS	http://www.fao.org/gtos/doc/pub55.pdf
Helsinki University	http://www.sciencedirect.com/science/article/pii/S0378112712007402
INRA Row Crop	http://www.sciencedirect.com/science/article/pii/S0168192310001206
LICOR LAI-2000/2200	http://envsupport.licor.com/docs/LAI-2200C_Instruction_Manual.pdf
Ryu, Nilson	http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.8023&rep=rep1&type=pdf http://www.cnr.berkeley.edu/biometlab/pdf/Ryu et al 2010 AgForMet clumping.pdf
UNECE	http://www.icp-forests.org/pdf/FINAL_Litter.pdf (http://icp-forests.net/page/icp-forests-manual)
VALERI http://w3.avignon.inra.fr/valeri/	http://research.eeescience.utoledo.edu/lees/papers_pdf/Weiss_2004_AFM.pdf http://w3.avignon.inra.fr/valeri/fic_htm/methodology/main.php http://w3.avignon.inra.fr/valeri/fic_htm/methodology/main.php